YMP/91-25

U.S. DEPARTMENT OF ENERGY

OCRWM

YUCCA MOUNTAIN

YUCCA MOUNTAIN SITE CHARACTERIZATION PROJECT

Testing Priorities at Yucca Mountain: Recommended Early Tests to Detect Potentially Unsuitable Conditions for a Nuclear Waste Repository

Volume I

Report of the Test Prioritization Task Force

March 1,1991

102.8

UNITED STATES DEPARTMENT OF ENERGY

Testing Priorities at Yucca Mountain: Recommended Early Tests To Detect Potentially Unsuitable Conditions for a Nuclear Waste Repository

**Report of the
Test Prioritization Task Force**

Volume I

Steven R. Mattson, chairman
Science Applications International Corporation

Bruce R. Judd
Decision Analysis Company

Scott R. Sinnock
Sandia National Laboratories

Dwight T. Hoxie
U.S. Geological Survey

March 1, 1991

Executive Summary

Project Objective and Analytic Approach

The objective of this study is to identify tests that could be conducted early during site characterization to detect the presence of potentially unsuitable site conditions for a nuclear-waste repository at Yucca Mountain, Nevada. The study analyzed 32 potential concerns (PCs) in detail, which were derived primarily from the potentially adverse and disqualifying conditions listed in 10 CFR Part 960 and the potentially adverse conditions in 10 CFR Part 60.122(c). The 32 PCs are listed in Table ES-1 together with brief definitions, reference numbers used to identify the PCs in the report, and short titles used to label the PC in subsequent figures. Complete definitions of the PCs can be found in Appendix C.

Table ES-1. List of potential concerns

Title of potential concern (PC)	PC number	Short title	Definition
Gas-flow radionuclide	1	Gas	Potential gas-phase release through the unsaturated zone to the accessible environment
Complex geology–Gaseous	2.2	CG-Gas	Incorrect prediction of gaseous-phase releases (by 10% of the baseline) because of expected or unexpected complexities in site geology or in the modeling of the site
Complex geology–Aqueous	2.1	CG-Aq	Incorrect prediction of aqueous-phase releases (by 10% of the baseline) because of expected or unexpected complexities in site geology or in the modeling of the site
Direct intrusion	H3	Intrusion	Direct intrusion of emplaced waste during drilling for water or economic resources
Expected GWTT<1000y	6	GWTT	Ground-water travel time (GWTT) less than 1,000 years along fastest path of likely and significant radionuclide transport
Oxidizing GW in host rock	4	Eh	Presence of chemically oxidizing ground water in repository host rock
Climate effect on Rn transp	5	Climate	Effects of possible future climate change on the site unsaturated-zone hydrologic system
Human intrusion effects-geohy.	H1	HI-Geo	Potential for future human activity to change hydrologic conditions at the site
Natural resources	H4	Nat res	Presence of potentially economically recoverable natural resources at the site
Perched water	8	Perched	Presence of perched water at or above the repository level
UO_2 Solubility	26	UO_2 sol	Solubility of UO_2 in ground water within the repository host rock
Past igneous act., site effects	16	Volcan	Potential effects of direct igneous intrusion into the repository
Reactive GW chem (EBS)	3	TDS	Presence of chemically reactive ground water in the repository host rock

Table ES-1. List of potential concerns *(continued)*

Usable water in CA: SZ	H8	Use H2O	Potential for future ground-water withdrawal to affect the site saturated-zone hydrologic system
Water-table rise: 200m	9	200m rise	Potential for water-table rise at the site > 200 meters due to future climatic or tectonic change
Therm/rad effects: corr. steam	24	Steam	Generation of corrosive steam in the repository environment due to heat released by emplaced waste
Future mining	H6	Fut mine	Potential effect of future mining near the site on unsaturated-zone hydrologic system at the site
Therm/rad effects: resat. flux	23	Resat flux	Potential effects on waste isolation due to thermally induced drying and subsequent resaturation near the repository
Past active tectonism (faulting)	13	Faulting	Potential effects on waste isolation due to future faulting within the repository
Rock & GW complex engr.	19	Cmpl engr	Presence of rock or geohydrologic conditions at the site requiring complex engineering methods to construct and close the repository
Geomorphic processes, past eros.	25	Erosion	Occurrence of future erosion rates at the site sufficient to affect waste isolation
Therm/rad effects: permeab. chg.	21	Perm chg	Potential for thermally induced permeability increases in the host rock near the repository
Tect eff. on reg. GW flow:UZ	11	Tect UZ	Potential effects of future tectonics on the site unsaturated-zone flow system
Past mining	H5	Old mine	Potential effects of past mining at the site on the unsaturated-zone flow system
Therm/rad effects: sorb. zeol.	22	Sorb Zeo	Potential thermally induced alteration of sorbing minerals near the repository
Tect eff. on reg. GW flow:SZ	12	Tect SZ	Potential effects of future tectonics on the site saturated-zone flow system
Water-table rise: 20m	10	20m rise	Potential for water-table rise at the site > 20 meters due to future climatic or tectonic change
Tectonic-induced lakes	15	Lakes	Possible occurrence of tectonic-induced lakes at the site
Past igneous act., CA effects	17	Volc CA	Potential effects of igneous activity near the site on waste isolation
Sorp/rock strength reduction	20	Rock str	Present rates of ongoing mineralogic change at the site
Rock cond. beyond RAT	18	Rock>RAT	Presence of rock conditions at the site requiring engineering methods beyond reasonably available technology
200m depth infeasible	7	200m depth	Inability to maintain a repository depth of at least 200 meters below land surface

Criteria for ***Evaluating*** *Tests*

This analysis is intended to answer the questions, "Which concerns have the greatest potential for rendering the site unsuitable with respect to possible postclosure radionuclide releases to the accessible environment?" and "Which tests are most likely to provide accurate detection of these concerns if they are present at the site?"

To answer these questions, it was necessary to develop specific definitions and quantitative measures for the PCs. These measures were introduced as surrogates for the PCs, and they are intended only for analyzing and recommending testing priorities. They are not intended to be criteria for evaluating suitability or unsuitability of the site, and to construe these definitions and measures as suitability criteria would be inappropriate.

Testing benefits in this analysis are expressed in terms of the consequences for the waste-isolation capabilities of the site. Tests are judged beneficial when they can detect potential concerns correctly and thereby allow decision makers to avoid the detrimental effects that may be caused by those concerns. Tests are not beneficial when they lead to "false alarms," that is, to indications that PCs are present when, in fact, they are not. Postclosure radionuclide release to the accessible environment over the next 10,000 years was used as the measure of all potentially detrimental effects on waste isolation and, therefore, on public health and safety.

This analysis found that very few of the potential concerns are of sufficient importance to merit early investigation for evaluating site suitability. This conclusion results either because the PC has a very small probability of occurring at the site or because its potential impact on waste isolation is deemed to be small. This conclusion does not mean that these concerns should not be investigated for other reasons, such as:

- Building scientific consensus about the evaluation of site suitability
- Gathering information for repository design and construction
- Providing ancillary information required for a license application
- Providing baseline data for long-duration performance-confirmation tests during and following repository construction.

A second conclusion of this analysis is that there is a high potential for false alarms from any test, regardless of the reason for testing. Decision makers need to be cognizant of the possibly significant consequences of false alarms as they plan the testing program.

Analytic Approach

The analysis was based on a five-step approach employing expert opinion and decision analysis techniques. Steps 1 and 2 identified and evaluated quantitatively the importance of the 32 potential concerns. More than 100 potential concerns were considered in Step 1, but they were screened and consolidated into the 32 PCs. Importance was defined as the product of expected consequences for waste isolation if the PC is present at the site with the probability that the PC is present. Consequences were measured in terms of the expected incremental increase in releases of radioactivity (in curies) to the accessible environment relative to expected baseline releases in the absence of any PCs. Expected releases were normalized by dividing the expected curies released by the Environmental Protection Agency (EPA) limits on releases (40 CFR Part 191.13).

The importance of each PC can be interpreted as the expected value of *perfect* information about its presence or absence at the site (i.e., 100-percent test accuracy). This sets an upper bound on the value of any practical testing activities that are aimed at detecting the presence of PCs. Step 2 screened the 32 PCs from Step 1 and identified 14 PCs that had the highest potential value for testing. These 14 PCs were then evaluated for test accuracy in Steps 3 through 5.

Step 3 identified the potential tests that could be used to detect the presence or absence of each of the 14 important PCs. "Tests," as defined in this report, may include one or more investigations, studies, or activities identified and described in the Site Characterization Plan. In some cases these tests were evaluated as a single package; in other cases progressive "levels" of test packages were evaluated. These levels generally progressed from less comprehensive to more comprehensive investigations. For example, a typical progression might include:

- Level 1. No new boreholes; use currently available data and non-surface-disturbing work
- Level 2. Limited surface-based drilling plus Level 1 data
- Level 3. Data from the Exploratory Shaft testing plus Levels 1 and 2 data.

Each test package for each level of testing was evaluated as a whole; no attempt was made to prioritize specific tests or activities within a test package.

Steps 4 and 5 assessed the accuracy of each of the packages of tests and evaluated their net benefits. Test accuracy was quantified for each package of tests using two conditional probabilities:

- The conditional probability of correctly "finding" the PC given that it is present
- The conditional probability of falsely "finding" the PC if, in fact, it is not present.

These are referred to as "true" and "false" positives, respectively, where "positive" denotes any test result that indicates that the PC is present. A "false positive" test outcome is also referred to as a "false alarm."

The net benefits of testing are expressed as the weighted difference between the detection benefits ("true positives") and false-alarm costs ("false positives"). Detection benefits are measured by the expected radionuclide release that could be avoided if the PC were present and detected by the test. It is assumed in this report that such releases could be avoided through mitigation or by abandoning the site, or by some other response to the detection. However, no specific actions were analyzed explicitly. False-alarm costs are measured analogously, by assuming that, given the result of the test, *unnecessary* action (e.g., costly mitigation or abandoning the site) would be taken in the false expectation of eliminating radionuclide releases. In the present analysis, both detection benefits and false-alarm costs are assumed to be *proportional to* the potential amount of radionuclide release that could be prevented by acting as if the PC were present. The evaluation assumes that action would be taken to avoid *all* incremental curies whenever the presence of a PC is detected.

Two value judgments by management are required to complete the analysis of test priorities. These include:

- The minimum detection benefits required to justify the dollar costs of testing to detect potentially unsuitable site conditions (i.e., whether the benefit of the information is worth its dollar cost)
- Benefit-cost weights, which are the relative values given to detection benefits and false-alarm costs when computing the net benefits of testing.

Ultimately, decision makers responsible for the site-characterization program must make these two value judgments, either explicitly or implicitly when they allocate funds to testing programs.

Analysis Results and Insights

Importance of Potential Concerns

Figure ES-1 displays the relative importance of the 32 PCs as determined in Step 2 of this analysis. The height of each vertical bar indicates the expected consequences

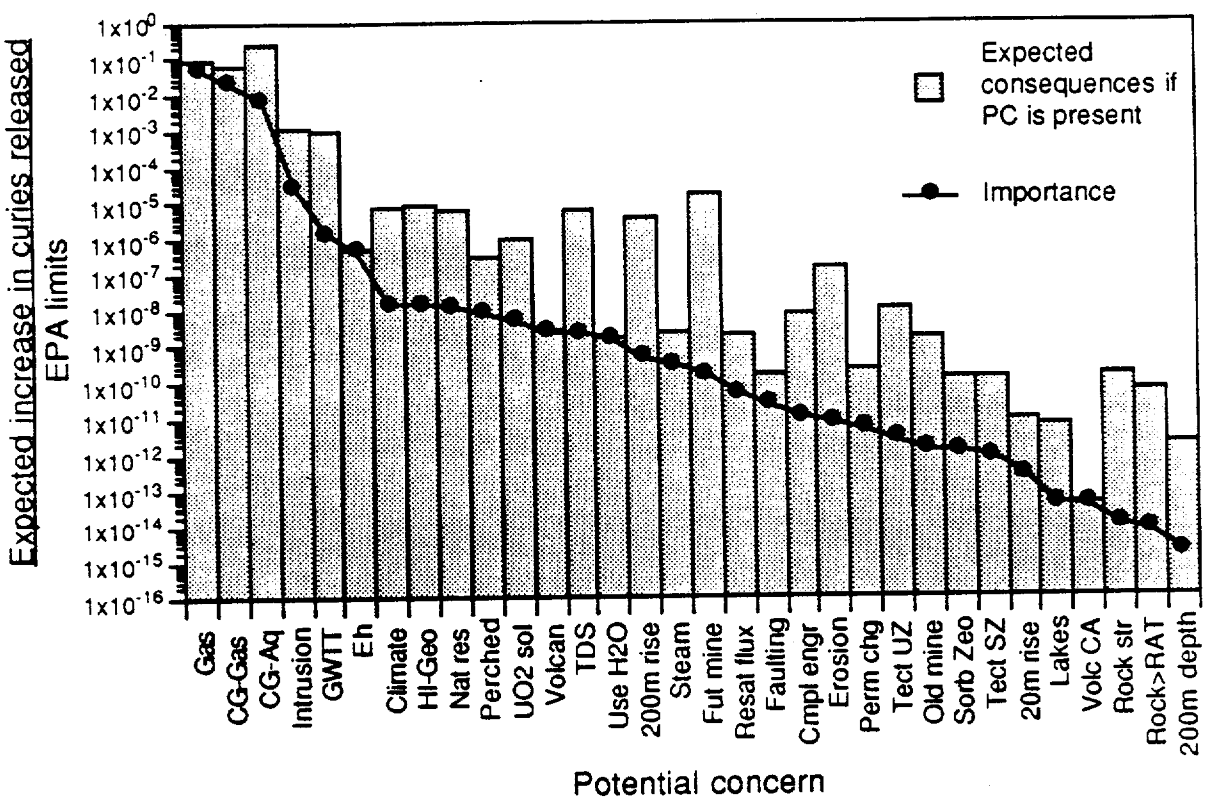

Figure ES-1. Importance and expected consequences of potential concerns. Vertical bars represent consequences *if* the potential concern (PC) is present, i.e., the increase in expected releases, normalized to the EPA limits. The dots and solid line represent the importance of the PC: the product of the probability that the PC is present and the expected consequences if it is present. A value of one on the vertical axis (top of the graph) corresponds to 700 excess cancer deaths over 10,000 years, or one excess death expected every 15 years (for a repository containing 70,000 metric tons of heavy metal). Abbreviations for PCs are found in Table ES-1.

(increase of normalized curies released) if each PC were present at the site. The dots connected by the solid line indicate the importance of each PC, which is the *expected* consequences of each PC weighted by the probability that the PC actually is present. (The line is provided as a visual aid and does not imply a functional relationship.)

As is apparent from the dotted line in Fig. ES-1, the importance of the 32 PCs spans a range of 14 orders of magnitude. Thus, on the basis of the expected effects on waste isolation, some of the PCs are much more important than others. The importance of a PC can be placed in a public health risk perspective by translating radionuclide releases over 10,000 years into expected excess cancer deaths for the same period. The relationship given by EPA in 40 CFR Part 191.13 allows one to equate the normalized radionuclide release of 1.0, shown at the top of the vertical scale in Fig. ES-1, to approximately 700 expected excess cancer deaths over 10,000 years for a repository containing 70,000 metric tons of heavy-metal waste. Using this conversion factor, the releases for the PCs shown in Fig. ES-1 range from a high of 39 to a low of 1.3×10^{-12} expected excess cancer deaths over 10,000 years (i.e., about one every 250 to 7×10^{15} years). The EPA limit on site performance for this repository allows 700 excess cancer deaths per 10,000 years (one every 14 years).

The set of PCs considered in this study were divided into three Importance Groups:

1. High relative importance: Three PCs that relate to releases of gas-phase radionuclides (specifically, carbon-14) and to complex site geology that affects gaseous and aqueous releases and could significantly complicate modeling site performance
2. Medium relative importance: Eleven PCs that relate to human intrusion, ground-water travel time, geochemical conditions in the host rock, perched water, and igneous activity at the site
3. Low relative importance: All 18 remaining PCs shown in Fig. ES-1

Because the PCs in Group 3 are unlikely to affect waste isolation, they were judged not to require further evaluation in this study. The PCs listed in Groups 1 and 2 were carried forward to the test-accuracy part of the analysis. Of these 14 PCs, four were combined to yield a final group of 10 PCs for which test accuracy was assessed. A total of 15 packages of tests for these 10 PCs were evaluated with regard to their accuracy for detecting the presence of a PC.

Test Priorities

As mentioned earlier, the net benefits of testing are defined in this report as the weighted difference between the detection benefits and false-alarm costs. Detection benefits are the expected radionuclide releases that could be avoided if the PC were both present and detected by the test. False-alarm costs are expressed as radionuclide releases *unnecessarily* avoided because of an erroneous belief that the PC is present.

Figure ES-2 is a plot of the detection benefits of tests versus the false-alarm costs. The left axis of the graph is scaled by the curies whose release to the accessible environment could potentially be avoided, and the corresponding right axis is scaled by excess cancer deaths potentially avoided in 10,000 years for a repository with 70,000 metric tons of heavy metal.

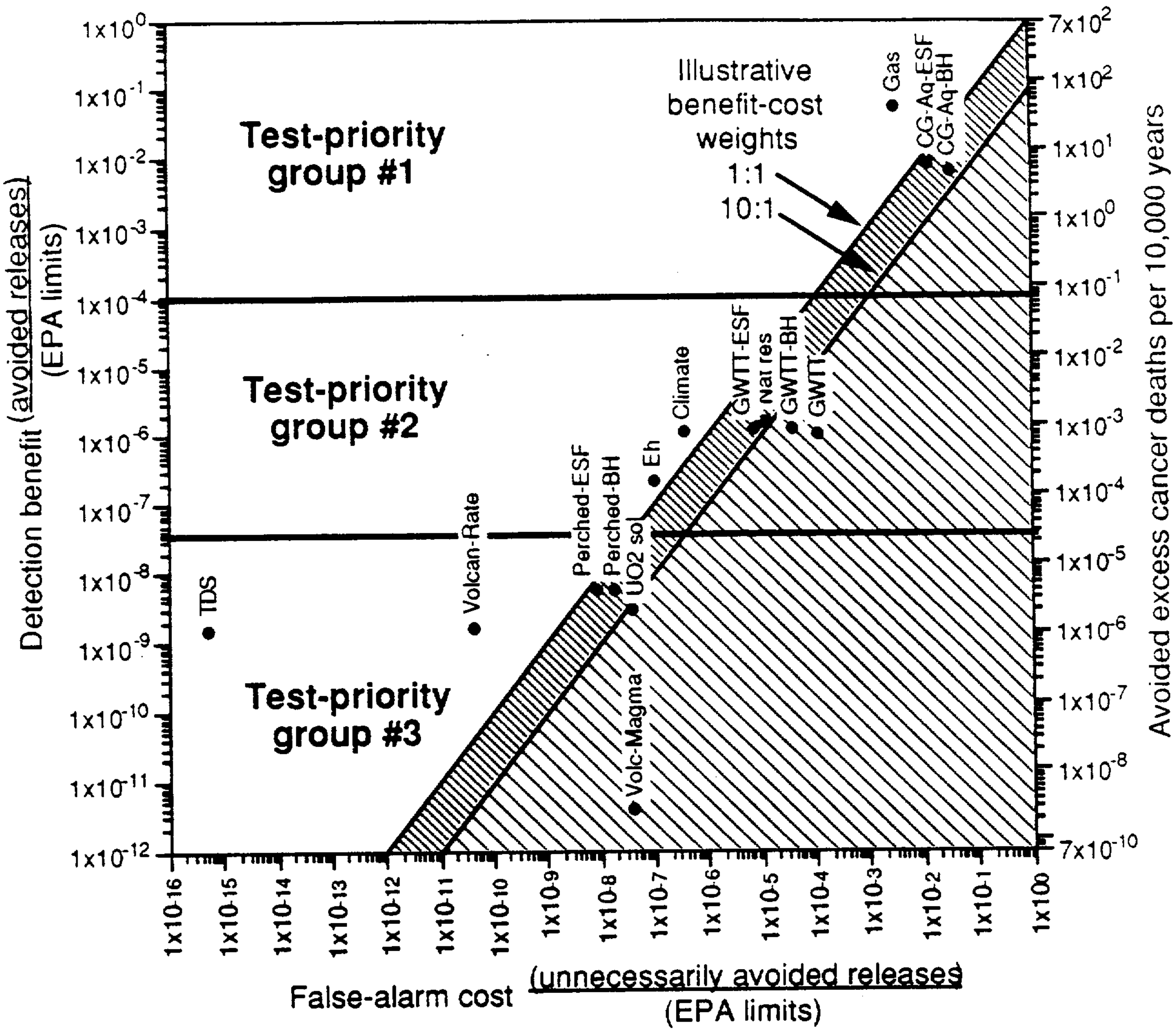

Figure ES-2. **Results of Phase I test prioritization.** This chart plots detection benefits and false-alarm costs for the 15 potential concerns (PCs) and their associated tests. Detection benefits and false-alarm costs are measured in terms of avoided curies released to the accessible environment and avoided excess cancer deaths over 10,000 years. Multiple test packages were evaluated for some PCs, for example ground-water travel time had three packages: GWTT, GWTT-BH, and GWTT-ESF. The abbreviations appended to GWTT are for borehole tests and Exploratory Shaft Facility tests, respectively. Two horizontal lines divide the chart into three test-priority groups, based on detection benefits. The detection benefits of tests in any particular group are roughly the same, and they are at least a factor of 100 different from detection benefits of tests in any other group. A decision maker chooses whether or not to conduct tests at each priority level based on a value judgment: whether the detection benefits at that level are large enough to justify the dollar cost of early testing for unsuitable site conditions. Within a test-priority group, one decides which tests to conduct based on a value judgment regarding "benefit-cost weights," which are used to weight detection benefits (vertical axis) and false-alarm costs (horizontal axis) when determining the net benefit of testing. Tests in the shaded areas have false-alarm costs that exceed their detection benefits and, therefore, should not be conducted solely to detect potentially unsuitable site conditions. If such tests are justified for other reasons, then one needs to recognize that false alarms are likely and one needs to develop a strategy for dealing with test results that may falsely indicate the presence of a PC. The two shaded areas in this figure are illustrative only. They represent hypothetical judgments that the value of an accurate detection is *equal to* a false detection (the darker shade) and *ten times* more valuable than a false detection (the lighter shade).

On the basis of the assessed detection benefits and false-alarm costs shown in Fig. ES-2, the PCs and their associated tests were prioritized into three test-priority groups. The groups are separated by at least two orders of magnitude in detection benefits.

1. High-priority tests
   Tests for gas-flow radionuclide transport above the repository and, depending on the value judgment on benefit-cost weights, the tests that address complex geology related to aqueous-phase radionuclide releases

2. Middle-priority tests
   Tests for climate changes, oxidation potential of water in the host-rock, and, depending on the benefit-cost weights, tests related to ground-water travel time, natural-resources, and direct-intrusion PCs.

3. Low-priority tests
   Tests related to reactive ground water in the host rock, rate of volcanism and, depending on the benefit-cost weights, perched water and UO_2 solubility.

The tests in Test-priority Groups #2 and #3 have low detection benefits, relative to those in Test-priority Group #1. A value judgment is required to determine if those benefits are sufficient to justify the dollar cost of testing to detect potentially unsuitable site conditions.

A second value judgment is needed to complete the prioritization of tests within each test-priority group. This is the relative importance of accurate detections versus false alarms, referred to earlier as the benefit-cost weights. The two diagonal lines in the figure represent two hypothetical value judgments about the relative weight of accurate detections to false alarms: 1:1 and 10:1. Tests represented by points in the shaded areas to the right and below the diagonal lines have false-alarm costs that exceed their detection benefits. Such tests cannot be justified solely for the purpose of detecting potentially unsuitable site conditions. Even if they are undertaken for "other reasons," one should bear in mind that they are very likely to yield a false alarm, and one should develop a strategy for dealing with positive test results that may falsely indicate the presence of a PC.

Summary of Conclusions

The major conclusions of this study are listed below:

1. *Importance of Potential Concerns*
From the perspective of potential effects on waste isolation, some of the PCs are much more important than others. Three PCs have greater potential contribution to radionuclide releases than all others by at least a factor of 200; these include "Gas flow radionuclide," "Complex geology-gaseous," and "Complex geology-aqueous." Among these three PCs, the highest expected contribution to curies released over 10,000 years is .06 times the EPA limits.

2. *Test Accuracy*
Test accuracy was assessed for 15 test packages associated with the ten most important PCs. Test accuracy ranged from 50 to 98 percent probability of detecting the PC if it is present. False-alarm probabilities ranged from nearly zero to 29 percent. Because the probabilities of true and false positive are coupled for a particular test (and one or the other can be made arbitrarily high), test accuracy alone is not a good measure for prioritizing tests. The probability that the PC is present and the consequences if it is present also need to be taken into account.

3. *Detection Benefits*
Detection benefits measure the expected contribution of a test for detecting a PC if it is present and thereby allows action to be taken to prevent the possible consequences of the PC. Detection benefits for the 15 evaluated tests ranged from .05 to 4×10^{-12} times the EPA release limits. Expressed as avoided cancer deaths, this is roughly one excess death avoided every 250 to 3×10^{12} years, respectively.

4. *False-Alarm Costs*
False-alarm costs also varied substantially: from .01 to 5×10^{-16} expected releases, or 8 to 3×10^{-13} cancer deaths, for which time and resources would have been expended unnecessarily, either for mitigation measures or for abandoning the site. For this reason some PCs may have false-alarm costs associated with early testing that may exceed their detection benefits.

5. *Testing priorities*
The tests of highest priority are those for gas flow (carbon-14 release) above the repository and, possibly, tests that address complex geology related to aqueous-phase radionuclide releases, depending on the value judgment on benefit-cost weights. These tests in Test-priority Group #1 have the potential to avoid one excess cancer death roughly every 280 to 2,500 years. The tests in Test-priority Group #2 contribute to avoiding only one excess cancer death every 10 million to 70 million years.

Tests for complex geology are worthwhile if the benefit-cost weighting is at least 10:1 (i.e., the curies avoided through early detection are worth at least 10 times more than the costs associated, with action taken to unnecessarily avoid releases due to false alarm).

6. Ground-Water Travel Time (GWTT) Sensitivity
The consequences of the PC regarding GWTT were related to cumulative releases for this study; whereas the requirement that GWTT be less than 1,000 years is a separate performance objective. If the consequences of violating the GWTT performance objective are set equal to the consequences of violating the EPA radionuclide release limits, then the detection benefits for GWTT tests increase by a factor of 600. This translates into a detection benefit of avoiding one excess cancer death over 20,000 years. However, the false-alarm costs increase proportionally, which makes the net benefits negative for tests to determine whether the GWTT is less than 1,000 years, unless relative benefit-cost weights greater than 100:1 are assigned.

7. Value Judgment on Minimum Detection Benefits
A judgment regarding the minimum level of detection benefit required to justify the dollar costs of testing is required before one can judge whether to conduct any tests in a particular test-priority group. Because there are only three distinct groups of tests, one needs only to choose among three minimum detection benefit levels. The minimum detection benefit in Test-priority Groups #1, #2, and #3 can be expressed as avoiding at least one excess cancer death every: 2,500 years, 70 million years, or 4 trillion years, respectively.

8. Benefit-Cost Weights
A second value judgment is needed to determine whether tests with high false-alarm costs can be justified. If the weights for detection benefits are judged to be equal to those for false-alarm costs, then the costs of many tests outweigh their benefits. This is because the tests are investigating unlikely and/or inconsequential PCs.

Recommendations

Based on the results of this analysis, the authors developed a set of recommendations in several topical areas, which include:

- Assessment of management value judgments
- Priorities for early tests
- Analyses of related issues
- Completion of the Phase II analysis
- Potential use of results in site-suitability determinations
- Further application of the approach to revise and update test priorities during site characterization.

1. Assess Management Value Judgments

Assessment of two types of value judgments by management personnel is required in order to set initial priorities on early testing to detect potentially unsuitable site conditions. The two types of value judgments are:

- Minimum detection benefits required to justify the dollar costs of testing to detect potentially unsuitable site conditions
- Relative benefit-cost weights of correct and false detection of PCs.

Once these judgments are made, Fig. ES-2 gives a clear indication of priorities for tests to detect PCs. The assessment of value judgments may be facilitated if, first, an analytic framework is developed to identify and related important factors to be considered in the assessment.

2. Set Priorities for Early Tests

Once the two value judgments are made, attention can be focused on deciding which specific tests should be conducted early during site characterization. Because the potential for releases is highest for gas-phase carbon-14 (C-14), this concern received highest priority in the evaluation. Although the assessment team identified tests that could be applied to gas-flow time above the repository and potential chemical retardation of C-14 transport, there is no testing program that is directed specifically to C-14 release and transport from the repository. Consequently, it is recommended that a *strategy* be developed for addressing potential C-14 releases. Although not addressed in this study, various other options may be available and should be considered when developing a strategy for C-14 releases. Some of these options include:

- Conducting site tests to evaluate the potential for rapid transport of C-14 to the accessible environment if it escapes from the waste containers
- Testing the waste form and cladding to determine the amount of C-14 expected in the rapid-release fraction
- Venting the waste before emplacement
- Reviewing regulatory requirements regarding C-14 releases and consideration of rule changes.

The tests for air flow (air permeability) and C-14 retardation were planned to support other objectives, primarily the characterization of hydrologic features of the unsaturated zone above the repository. Such tests could be given high priority if site testing is a part of the "C-14 strategy." However, it is recommended that a testing strategy explicitly focused on C-14 transport factors be developed before assigning high priority to the currently-identified applicable tests.

Tests for complex geology could be assigned high priority, depending on a clarification of the relation between complex geology and modeling accuracy and on the management value judgment about benefit-cost weights (Recommendation 1 above). Specific tests for early evaluation of complex geology include:

- Vertical borehole investigation of the Ghost Dance and Solitario Canyon faults

- Potentiometric-level evaluation to investigate the steep-gradient zone north of the site
- One to three boreholes from the systematic drilling program that are independent of specific features in order to provide areal control.

These tests were evaluated together, as a package. No prioritization was evaluated or implied for specific tests in the package.

This analysis does not support priority testing for other PCs for the purpose of early detection of potentially unsuitable site conditions. While there may be other reasons for giving high priority to such tests, one needs to be aware of the relatively high likelihood of these tests to yield *false* indications of unsuitable site conditions.

3. Analyze Preclosure and other Site-Suitability Issues
The authors recommend expanding the scope of the analysis to address preclosure or other site-suitability issues not addressed by this analysis. There may be good reasons among those issues to justify early testing of the site (e.g., seismic concerns related to preclosure operations).

4. Complete Phase II Assessment and Analysis
Several possible extensions of the assessment and analysis should be considered in Phase II. The number and diversity of workshop participants could be expanded to include a broader range of experts on individual PCs, possibly including experts external to the current program. The criteria for evaluating the importance of PCs could be expanded from the current postclosure total-system-performance criteria to include: preclosure health and safety, ease and cost of construction, environmental, socioeconomic, and transportation impacts, and postclosure subsystem performance. Similarly, the criteria for prioritizing tests could be expanded to include the dollar cost of tests and measures related to the "other reasons" for testing listed earlier.

In addition, expansion of the assessment and modeling should be considered. A dominant factor influencing the conclusions of this analysis is the expected consequences for waste isolation if potential concerns are present. These consequences are based on expert judgment and were difficult to assess. A simplified total-system-performance model for calculating those consequences would enhance the credibility, clarity, defensibility, and future utility of this analytic approach. Further, the inputs to a total-system-performance model would be assessed at a lower level of detail, compared to Phase I assessments. This would make the assessments easier for the experts, especially in cases where several interrelated factors were considered simultaneously in the Phase I assessments. The authors recommend that managers consider and decide which of these expansions of the analysis will most enhance its usefulness as a management tool.

5. Potential use of Results in Site-suitability Determinations
The analytic method, assessments, and numerical results of this analysis can provide useful information and methods to the process being considered for early evaluation of site suitability. For example, the importance and testing assessments

yield insight on what might be learned in the initial phases of testing. This information bears directly on the "lower-level" and "higher-level" findings required by 10 CFR Part 960. However, to address suitability issues comprehensively, the scope of the analysis would need to be broadened, as discussed above, to incorporate preclosure and other related issues. The method developed here for prioritizing testing embodies a simplified analysis of three alternatives: continue testing, apply for a license, or abandon the site. This structure is compatible with and directly applicable to factors in site-suitability decisions that consider the net benefits of these alternatives.

6. Apply Method to Reprioritize Tests During Site Characterization
The analytic method used in Phase I can be extended and reapplied at any point during site characterization. The Phase II model would enhance such applications, but the procedure would be similar with or without the Phase II model. Phase I established a foundation for future assessment and analysis, and only changes to that foundation will be required in future applications.

Similarly, the assessments obtained in this analysis can be used to update assessments of the probability that various PCs are present, given the results of early testing. This is especially important if tests are conducted that have high probabilities of false alarms, as do many of the tests analyzed in this report.

In fact, an important issue for decision makers to face is, "How should one treat results from tests that are conducted for reasons unrelated to early determination of potentially unsuitable site conditions but whose results show that potential concerns may be present." According to the assessments in this study, these test results may well be false alarms. In summary, the method developed here provides a systematic and defensible approach that could be used for updating assessments with new information, drawing inferences, and making suitability or testing decisions based on test results.

Volume I
Table of Contents

List of Figures

List of Tables

Acknowledgments

This report represents a collaborative effort by roughly 100 people, each of whom made a significant commitment toward developing a carefully and logically reasoned set of priorities among proposed scientific tests at Yucca Mountain. The largest group of participants—the scientific experts—graciously contributed their time and expertise, often in multiple, extended sessions that competed or conflicted with other high priority demands on their time. These experts also assisted in implementing and refining this new approach to setting priorities and willingly provided expert judgments in a quantitative form that was often different from their normal mode of communication. Such cooperation provides strong evidence of their commitment to quality decision making regarding the Yucca Mountain site.

The task force is also deeply indebted to the contributions of many other essential players. J. Russell Dyer, the Department of Energy's (DOE's) supervisor of and customer for this effort, provided continuing, positive support through many project complications, delays, and organizational hurdles. Also strong in their support of the effort were DOE Yucca Mountain Project Office managers, including David C. Dobson, Maxwell B. Blanchard, and Carl P. Gertz. DOE Headquarters staff, including Jeffrey K. Kimball and Scott G. van Camp also contributed to the early stages of this effort.

William E. Wilson, recently retired from the U.S. Geological Survey, was one of the original members of the Core Team. Bill generously contributed his wisdom to the first half of the project, along with patient explanations to the Core Team's neophyte member about, for example, geohydrology, hydrogeology, and all science and lore in-between.

Timothy G. Barbour, SAIC, also was an early member of the Core Team and provided the output from the PARATRAC data base that served as the starting point for the testing workshops.

John Lathrop, of Strategic Insights, joined the Core Team at a critical point and provided valuable assistance developing the assessment strategies and graphic aids for the assessment workshops. His brief, but intense contribution provided the motive force that propelled several workshops toward their intended goal.

Several other members of the team also made many contributions that increased the overall quality of the effort. In particular, we thank Jeffrey D. Weaver, August C. Matthusen, Martha W. Pendleton, Robert P. Gamble, Sandra J. Trillo, Patricia J. Thompson, and Suzannah M. Marek for assisting in numerous ways, raising the quality of our product, and making the task an enjoyable one.

Finally, the team is indebted to its creator, Jean L. Younker. Her motivation, insight, and savvy were essential to the health of the project and its completion. However, it is the authors who bear full responsibility for any calculational errors or inaccurate representations of expert judgment that occur in this report *in spite of* her wisdom and guidance.

1. Introduction

Background

The Test Prioritization Task (TPT) was established by the Department of Energy (DOE) to assist in prioritizing the scientific investigations and studies planned as part of the evaluation of Yucca Mountain, Nevada, as a potential site for a mined, geologic repository for permanent disposal of high-level nuclear waste. The Task was initiated in January 1990 specifically to identify near-term tests that could provide early detection of potentially unsuitable conditions at the site. This was in response to a 1989 policy decision by DOE "...to focus its near-term scientific investigations...specifically at evaluating whether the site has any feature that would indicate that it is not suitable as a potential repository site." (DOE, 1989) The task was conducted by a Core Team, which comprised the authors of this report and other members listed in Appendix B. Assessments were made in a series of workshops; the participants in those workshops were referred to formally as the Integration Team.

In the course of investigating and characterizing the Yucca Mountain site, there will be many reasons for conducting individual tests, such as:

- Evaluation of site suitability.
- Building scientific and public confidence in the evaluation of site suitability
- Gathering information for repository design and construction
- Providing ancillary information required for a license application
- Initiating long-duration performance-confirmation tests during and following repository construction.

Although all of these reasons are amenable to the type of analysis described in this report, they are *not* all considered in this report. Rather, the analysis here is focused on prioritizing tests that could be used *during the early stages of site characterization to detect features or conditions that might indicate that the site is not suitable for the construction and operation of a repository*. Clearly, this is but one of many important considerations for DOE, and the results of the study provide some, but not all, of the key inputs to test-prioritization decisions.

The analysis method developed for this task provides DOE with a tool for prioritizing tests now and for periodically reassessing priorities as site characterization proceeds. This approach builds on the Site Characterization Plan (DOE, 1988), by providing explicit quantification and evaluation of the essential considerations needed for setting priorities among tests that could detect potentially unsuitable site conditions. These considerations include:

- The likelihood that specific unsuitable conditions or surrogate indicators of those conditions are present at the site
- The estimated consequences (releases of radioactivity) if those conditions or indicators are present but not detected

- The accuracy of tests for detecting those conditions or indicators
- The likelihood and consequences of "false alarms" (i.e., erroneous detection of a condition or indicator when, in fact, it is not present).

The same approach can be used to assist in decision making related to site suitability as test results are obtained. This approach, for example, could support interpretation of those results and decisions about whether management should continue, change, or eliminate related tests in light of new test results.

Scope

Originally, the scope of this study was restricted to tests included within the planned surface-based testing program. The study was expanded in September 1990 to include prioritization of other tests, such as Exploratory Shaft tests, analog research, etc. However, the primary focus of the TPT has remained on tests (surface-based or otherwise) that could detect adverse conditions early during the site characterization program and could directly influence management decisions about site suitability.

The analysis in this report is based on the use of performance criteria to identify and evaluate tests that could detect site features or conditions that affect the postclosure waste-isolation capabilities of the site to an extent that would raise serious questions about the site's ability to meet applicable performance objectives. In particular, this study used the Environmental Protection Agency's (EPA's) radionuclide release limits set forth in 40 CFR Part 191 as proxies for all applicable performance criteria. This regulation sets limits on radionuclide releases from the repository to the accessible environment.

Those site features or conditions that may adversely affect waste isolation are referred to as "unsuitable" subsequently in this report. However, each condition discussed in this report potentially affects waste isolation in different ways or to varying degrees. Therefore a critical feature of the analytic approach is to quantify the magnitude of the potential effects on waste isolation and the influence of those effects on test priorities. Whether these effects are unacceptable, and therefore render the site unsuitable, is a judgment that will be made prior to or during preparation of an application to license a repository at the site. This judgment will be based on the regulatory requirement that "reasonable assurance" be demonstrated and that limits on the cumulative release of radionuclides, established by the EPA in 40 CFR Part 191, will not be exceeded.

Although this study was performance-based, no explicit performance assessment calculations or evaluations were undertaken as a part of this study. However, the results of available performance assessment calculations were considered where appropriate. A second part of this study is planned (Phase II) that would use a total-system-performance model to support, refine, and possibly change the test-prioritization results produced by the present study. It is important that *this study not be construed as an analysis of the suitability or unsuitability of the site,* even though there is much discussion here about effects on waste isolation, site performance, the compliance with or violation of criteria, and site suitability or unsuitability. Rather, this analysis assesses the *uncertainty* about the effects of various factors on site suitability and the degree to which various tests could resolve or reduce that uncertainty. The type of analysis conducted here, specifically for test prioritization, is different from the type that would be necessary to judge the site as "suitable," "licensable," "ready to accept waste," or "ready for closure, etc."

Experts with diverse scientific, engineering, and regulatory backgrounds were recruited to provide input for this analysis. These experts included representatives from many organizations intimately involved in planning or conducting site characterization and evaluation activities (the U.S. Geological Survey, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, Sandia National Laboratories, Lawrence Berkeley Laboratory and Pacific Northwest Laboratory); universities (including University of California at Berkeley); and consulting organizations (including Science Applications International Corporation, Roy F. Weston Inc., and Disposal Safety Inc.). These sources are listed in Volume II, Appendix B.

Two other projects, the Calico Hills Risk/Benefit Analysis and the Exploratory Shaft Alternatives study, were conducted concurrently with the TPT effort and had related objectives. These three tasks were coordinated to ensure consistency among the analytic approaches, where appropriate, and consistency of assumptions and data. In many cases this consistency emerged from sharing common participants in the multiple studies.

However, all three tasks operated under significant constraints on time and the availability of experts. Therefore, the TPT adopted an iterative and phased approach to prioritization, as explained in Chapter 2. This phased approach produces an initial set of recommended test priorities on an expedited schedule, but its full implementation will involve multiple iterations of the analysis. The results of subsequent, more refined analyses may revise the priorities established in this initial phase of the test-prioritization effort.

Type of Analysis and Results

The data assessments and analyses described in this report are based on decision analysis, which is a systematic and quantitative approach for guiding decision making in complex situations subject to high uncertainty where unaided intuition is inadequate. Setting test priorities is a classic case where unaided intuition can lead one astray because of the need to consider a large number of factors simultaneously. These factors include:

- The selection of conditions to be investigated by testing
- The accuracy of these tests in detecting the conditions
- The range of possible results of testing

- How those results will be incorporated in decision making once they are available.

Decision analysis provides logical and defensible techniques for quantifying how testing will resolve inherent uncertainty. These include methods for assessing the degree of uncertainty prior to testing, the accuracy of tests, and the residual uncertainty after test results have been obtained. Quantifying expert judgment on each of these types of factors is essential for logical analysis of the benefits of resolving uncertainties regarding early detection of potentially unsuitable site features or conditions. Therefore, probability assessment and probabilistic analysis are essential features of this approach.

The results of the analyses in this report are lists of prioritized packages of tests. (The "test packages" are highly aggregated sets of individual tests as will be explained later and are hereafter referred to, simply, as "tests.") Much more important than the numbers and the lists themselves, however, are the *insights* gained regarding why certain tests may be ranked as high, medium, or low priority. Therefore, Volume I of this report emphasizes these insights, rather than input data, mathematical equations, or analytic detail. Volume II provides the input data, analysis methods, results, and other technical information.

Tests receive high priority in this analysis if they are likely to detect potentially unsuitable site conditions. This occurs when several of the following conditions obtain:

- The features or conditions are likely to be present
- The proposed test has high likelihood of detecting the conditions if they are present and low probability of "detecting" them incorrectly if they are not present
- There may be a significant detriment to waste isolation if such features or conditions are present but are not detected.

However, a strong conclusion of this report is that many of the tests that have been suggested as candidates for early detection of unsuitable site conditions did not receive high priority in this analysis. Such results occur for one or more of the following reasons:

- The features or conditions addressed by the tests are unlikely to be present, based on what is known today about the Yucca Mountain site
- If they are present, these features are likely to be inconsequential, with respect to waste isolation
- Available testing technology may be inadequate for detecting or evaluating existing unsuitable features
- The tests are likely to conclude erroneously that unsuitable features or conditions are present
- The tests are too expensive relative to the value of information they provide.

It is important for decision makers to keep these reasons in mind while reviewing the results of this analysis, and, thereby to understand why many tests were accorded low priority for the early detection of potentially unsuitable conditions. Chapter 3 presents several different graphic displays that illustrate these reasons. It is also important to remember that there are many other reasons for testing, some of which were listed at the beginning of this chapter. Thus, the tests that do not receive high priority for early detection of potentially unsuitable conditions may well receive high priority for other reasons.

Organization of the Report

The report is organized in two volumes: Volume I is the main body of the text and summarizes the approach, results, and conclusions. Volume II contains a set of appendices that provide technical detail.

Chapter 2 in Volume I summarizes the general analytic approach and methods for assessing quantitative inputs for the analysis. Chapter 3 describes the analysis itself: inputs, intermediate results, and final results. Chapter 4 provides conclusions regarding testing priorities and recommendations for subsequent phases of the analysis.

The appendices in Volume II cover several topics in more detail. Appendix A provides additional detail on the sources of potentially unsuitable features and conditions analyzed in this report. That appendix also discusses their relationship to the regulations 10 CFR Part 60 and 10 CFR Part 960. Appendix B details the Test Prioritization Task History, including lists of participants in assessment workshops, a bibliography, and lists of related project correspondence. Appendices C and D list the assessments provided by participants in two sets of workshops. These appendices also summarize some of the discussion in the workshops regarding the assessments.

2. Analytic Approach

In order to achieve the objective of prioritizing tests that could be conducted early and that could detect potentially unsuitable site conditions, an analytic approach and assessment process were developed to produce a quantitative basis for setting priorities. As mentioned in the introduction, a phased approach was adopted in order to meet the DOE's needs in a manner consistent with available time and resources.

This chapter begins with a general description of the phased approach, and the five-step analytic method used to implement Phase I of the prioritization effort is then explained. The chapter concludes with a summary of the assessment process used to quantify expert judgments, especially those regarding the uncertainties that might be resolved through testing.

Phased Approach

Two compatible decision analysis methods have been developed, which can be considered Phase I and Phase II of the same test prioritization effort. Phase I is intended to be preliminary to Phase II and, subsequently, to guide the Phase II effort. Phases I and II are scheduled to provide results that can be used in setting priorities in 1991 and 1992, respectively, for the early phases of the site-characterization program.

The Phase I analysis, which has been completed and is reported here, is based on available information and expert judgments about the following factors:

- The features and conditions characteristic of a potentially unsuitable site, which are designated potential concerns (PCs) in this report; these include the potentially adverse and disqualifying conditions listed in 10 CFR Part 960 and the potentially adverse conditions in 10 CFR Part 60.122(c)
- The probability that these PCs are present at the Yucca Mountain site
- The extent to which these PCs affect waste isolation
- The accuracy of tests for detecting the presence of these PCs
- The relative consequences of accurately finding PCs that are present, or erroneously "finding" PCs that, in fact, are not present.

The information and quantitative judgments about these factors were analyzed using a personal computer "spreadsheet" model to produce the tables and graphs in this report. The methods developed during this Phase I analysis can be extended to evaluate the dollar costs of tests or "other factors" that are important for prioritizing tests but that do not directly address the site's suitability for isolating waste. These other factors, however, were not assessed during this Phase I study and are not considered in this report.

One of the most difficult judgments made during Phase I was the extent to which PCs are likely to affect waste isolation. Answering this question may require consideration of many complex and interrelated factors. Due to time constraints, the questions had to be posed at a highly aggregated level, for example, the effect of a 20-meter water-table rise on total-system-performance. Although the experts who participated in this study were able to make informed judgments about these effects, the range of opinion frequently varied greatly. A few of the analyses were found to be sensitive to the range of judgments concerning the potential effects of PCs on waste isolation.

Phase II is designed to refine such judgments, using simplified calculations in a total-system-performance (TSP) model. This will allow expert judgments to be assessed at a lower level of detail, for example, the effect of a 20-meter water-table rise on hydraulic gradients and other hydrologic properties. These assessed changes will then be analyzed by the model to predict changes in ground-water flow time and radionuclide released to the accessible environment.

Development of the Phase II model proceeded concurrently with the Phase I effort, but the model was not available to meet the Phase I prioritization deadline (see historical summary in Appendix B). Thus, the subjective assessments of the effects of PCs on waste isolation serve as an initial application, until more detailed analyses can be performed. The planned Phase II effort will also be implemented on a computer spreadsheet, but it will be based on the results obtained from the TSP model. Although application of the Phase II performance model will not yield definitive predictions of the effects of features or conditions on system performance, it is the opinion of the Core Team that such a model will provide additional insight and significantly improve the defensibility of estimates made by the experts during Phase I.

Phase II could also expand the criteria for evaluating the importance of PCs from the current postclosure total-system-performance criteria to include: preclosure health and safety, ease and cost of construction, environmental, socioeconomic, and transportation impacts, and postclosure subsystem performance. Similarly, the criteria for prioritizing tests could be expanded to include the dollar cost of tests and measures related to the "other reasons" for testing, where were listed in the introduction.

Phase II also will include assessments of information from a broader range of experts than was utilized in Phase I. The Phase II TSP model will be completed in time to be used as a management tool for prioritizing tests in 1992 and serve as input to the evaluation of site suitability. In summary, Phase II will be more robust than Phase I in the following respects:

- Broader range of expert input, possibly including experts external to the current program
- Easier, disaggregated assessments of model inputs

- More defensible results, especially the analysis of the effects of PCs on waste isolation
- Greater capability for analyzing simultaneously the effects of various PCs on waste isolation
- Better able to determine the sensitivity of waste isolation capability to PCs and the results of testing.

The next section describes in more detail the steps in the Phase I and Phase II analytic methods.

Steps in the Analytic Method

Phase I is based primarily on expert judgment. Phase II is supplemented with more robust analytic models. However, both phases follow the same general analytic approach for prioritizing tests.

The general approach is implemented in five steps, which are illustrated in Fig. 2-1. The concept of each step is described very briefly in the paragraphs below. Chapter 3 provides the details on how the steps were implemented and discusses the assessments and analysis results.

Step 1. List Potential Concerns (PCs).
Potential concerns (PCs) are features or conditions that may cause the site to be unsuitable and, therefore, are possible "targets" of early testing. Most of the concerns on the list are derived from the potentially adverse conditions (PACs) set forth in 10 CFR Part 60 and 10 CFR Part 960 and the disqualifying conditions in 10 CFR Part 960.

In Step 1 the Core Team defined each PC in terms that are specific enough to allow the assessments in subsequent steps. The "definitions" include quantitative "measures" for judging the magnitude of the concern and "assessment thresholds" that specify the level at which the concern might be expected to have consequences with respect to waste isolation.

Step 2. Assess and Rank the "Importance" of PCs
This is a "screening" step used to avoid expending unnecessary effort in Steps 3 through 5 on low-impact PCs. This step ranks the PCs in terms of their importance to waste isolation (and hence their association with conditions that could cause the site to be unsuitable). Those PCs with low assessed importance to waste isolation are unlikely to justify early testing and are not carried forward to Steps 3 through 5.

This screening step produces an ordered list of PCs, ranked according to their assessed importance to postclosure waste isolation. The ordered PCs were divided into high-, medium- and low-ranking PCs in the discussion of Step 2 in Chapter 3. Low-ranking PCs were not considered further, because either they had very low probability of being present (hence it is unlikely that they would be found by testing)

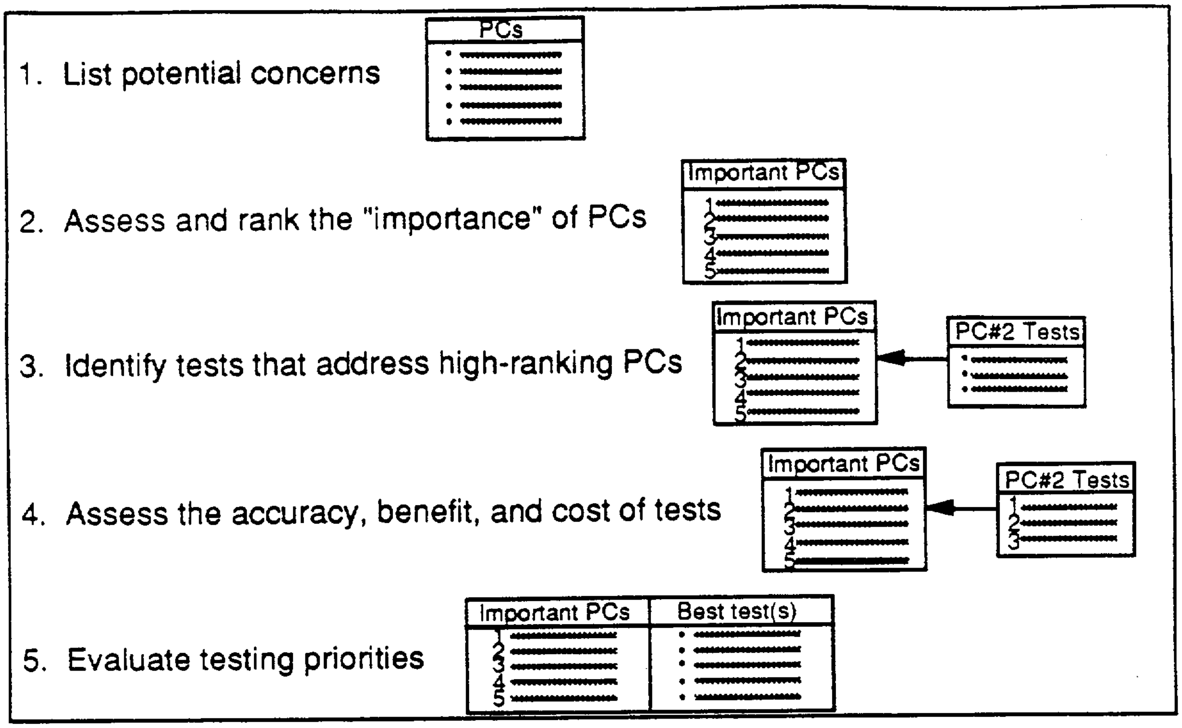

Figure 2-1. Five-step analytic approach to setting priorities. The approach gives priority to tests that can accurately detect potentially unsuitable site conditions early during site characterization. The potential concerns (PCs) being investigated are based on the potentially adverse conditions set forth in 10 CFR Part 60 and 10 CFR Part 960 and postclosure disqualifying conditions identified in 10 CFR Part 960. Step 1 identifies the concerns, and Step 2 ranks them according to their probability of occurrence and consequences if they are present at Yucca Mountain. Step 3 identifies potential tests for each specific PC, based primarily on the Site Characterization Plan. Step 4 assesses the accuracy, benefits, and some costs of those tests and identifies the best test to investigate each PC; Step 5 then establishes an overall set of priorities for early testing to detect potentially unsuitable conditions.

or they were not very detrimental to performance (hence finding them with testing would be of minimal value).

Step 3. Identify Tests that Address High-ranking PCs

"Tests," as defined in this report, may include one or more investigations, studies, activities, or sub-activities, as identified and described in the Site Characterization Plan (SCP). An initial list of tests for each potential concern was compiled from the PARATRAC database in order to link tests with PCs. The initial list was refined by the workshop experts familiar with the testing program for each PC. Other tests not found in the SCP were specifically solicited from participants in the assessment workshops.

In most cases individual tests were grouped into "suites" or "packages" of tests related to a particular PC. The evaluations of test accuracy in Step 4 were then conducted for entire packages of tests. In some cases, three levels of testing packages were evaluated. For example:

- Level 1. No new boreholes; use currently available data and non-surface-disturbing work
- Level 2. Limited surface-based drilling plus Level 1 data
- Level 3. Data from the Exploratory Shaft testing plus Levels 1 and 2 data.

Step 4. Assess the Accuracy, Benefit, and Cost of Tests

The accuracy of a test is measured here using two probabilities:

- The probability that the test "finds" the PC if the PC is present
- The probability that the test erroneously "finds" the PC when the PC is not present.

These are referred to as "true" and "false" positives, respectively, "positive" denotes any test result that indicates that the PC is present. A "false positive" test outcome is also referred to as a "false alarm."

The benefits of conducting a test that could detect a PC are referred to as "detection benefits." These are quantified using the importance of the PC from Step 2 and the probability that the test finds the PC when it is present.

Two types of costs are considered in this analysis. The primary cost discussed is the "false-alarm cost," that is, the impacts of erroneous detection of PCs that are not present at the Yucca Mountain site. The other cost, which is treated only briefly at the conclusion of the report, is the dollar cost of conducting the test.

The weighted difference between detection benefits and false-alarm costs is interpreted here as the "net benefit of testing." (The weights are a management judgment.) Net benefit is the criterion used to rank tests and to identify a single "best" package of tests for each high-ranked PC. The next step compares the benefits and costs of testing for all PCs and identifies those tests with the greatest ability to detect the presence of potentially unsuitable site conditions.

Step 5. Evaluate Testing Priorities.

After the completion of the assessments in Steps 1–4, and the identification of the best test for each PC in Step 4, an overall ranking of the tests is established. Results are in the form: "Given the kinds of tests that are available, early investigation of potential concerns A, B, and C have the highest ranking with respect to detecting potentially unsuitable site conditions."

Step 5 clearly depends upon management value judgments, such as the relative weights for detection benefits and false-alarm costs. (This report uses the phrase "value judgment" to represent a statement that goes beyond purely technical considerations and reflects preferences of the individual making the judgment.) Although such judgments were not assessed during Phase I analysis, the results of the analysis are presented on a chart that allows the implications of different value judgments to be seen. These management judgments are explained in detail in Chapter 3.

Step 5 also examines the question of the "absolute" benefits of testing. Specifically, the analysis considers whether the detection benefits of testing are sufficient to justify the costs of changes to the site characterization program that may be needed in order to conduct particular tests early.

Strengths and Weaknesses
This five-step approach has several advantages and limitations. Its primary strength is that it identifies, in a systematic and defensible way, the potential concerns that should be investigated early in order to detect potentially unsuitable site conditions. It also accounts for the current state of knowledge, including the uncertainty associated with each potential concern and how well the uncertainties could be reduced through testing. Furthermore, the process of assessing these factors produces many insights into the merits of alternative testing strategies.

This approach also has important limitations, although, in the Core Team's opinion, these are not severe enough to invalidate the insights it generates:

- The detection benefits and false-alarm costs of testing considered in this approach are associated only with detecting potentially unsuitable site conditions. There may be other important reasons for testing, however.

- The detection benefits and false-alarm costs of testing in this approach are proportional to the importance of each potential concern. These are very difficult to estimate directly without the aid of a total-system-performance (TSP) model, which provides strong motivation for refining the priority list using the TSP model in Phase II of the study.

- While this approach explicitly considers how different tests can provide information about a single PC, it does not take into account how a single set of tests might contribute to the investigation of several PCs. Consequently, computing benefits of testing for each PC individually could miss the aggregate benefits of tests that address several PCs.

- The approach does not explicitly consider probabilistic dependence between PCs, for example when the conditional probability of one PC goes up if a related PC is known to be present. This effect is currently handled by coupling the PCs in their definition and treating them as a single PC.

- Scheduling and budgeting procedures were not taken into account and may render some "high priority" tests impractical to field early in the site-characterization program.

Assessment Method

Because the Phase I analysis is based on expert judgment, it is essential to have a systematic, unbiased assessment process that provides quality input to the analysis and, ultimately, to decision makers. An efficient and effective process evolved, founded on elicitation techniques commonly used in decision analysis. The process, however, was modified as needed to meet specific requirements of this analysis.

Most of the expert judgments involved assessments of probabilities. There is extensive literature on probability-assessment techniques, as well as a set of professional techniques for obtaining unbiased probability judgments. Theoretical foundations of subjective probability were set forth by Bayes in 1763 (republished, 1958), de Finetti (1937), Savage (1954), and Raiffa and Schlaifer (1962) and have been expounded upon by many others since. Tversky and Kahneman (1974) identify several motivational and cognitive biases inherent in subjective probability assessments, and systematic procedures have evolved in the professional practice of decision analysis (Spetzler and Staël von Holstein, 1975; McNamee and Celona, 1987; Bonano et al., 1990; Apostolakis, 1990). There are also many publications on specific private and public-sector decision analyses that rely heavily on subjective probability assessment: for example, Howard and Matheson (1984); Keeney (1980); Merkhofer (1987); Judd and Weissenberger (1982); Brown (1987).

The assessment process developed for this study involved several features that are described below. These features relate to the workshop format, selecting participants, avoiding biases in probability assessment, casting ballots, aggregating individual probability assessments, and documenting the proceedings.

Workshop Format

Assessment sessions were conducted, for the most part, in off-site workshops of two- to three-day duration. Each had a clearly defined agenda, a facilitator to set the pace and conduct assessments, a recorder to document the proceedings, and participants who were experts in the topical areas addressed.

Two types of workshops were held:

- Importance Workshops for assessing the importance of the 32 potential concerns (Step 2 in the analytic method). The Importance Workshops comprised two sessions, lasting two and three days, respectively.
- Testing Workshops for assessing the accuracy of testing for the 13 highest-ranked PCs (Step 4). The Testing Workshops comprised three sessions, lasting two, three, and two days, respectively.

There were 9 participants in the Importance Workshops and 17 in the Testing Workshops. The workshop dates, documentation, and participants are listed in Volume II, Appendix B.

On the first day of each set of workshops, participants were briefed on the workshop purpose and the methods of analysis. After demonstrations of the assessment methods, participants were allowed time to discuss the approach, modify it where necessary, and agree upon a method for arriving at "consensus" on quantitative inputs to the analysis. The remaining days in the workshops were devoted to assessing the planned information. Generally, the first assessments required much more time than subsequent assessments (e.g., as much as one-half day for one assessment). It took a significant period of time for participants to become comfortable with providing the required information and also for a team ethos to develop.

The workshops were conducted in separate two- or three-day sessions to ease the travel burden for participants (i.e., shorter periods away from the office and their normal responsibilities). This also provided time for "homework" and consultation with other experts between sessions.

Workshop Participants

All participants in the assessment workshops were formally "members of the Integration Team," as described in the management plan (DOE, Jan. 1990) for the TPT. For simplicity, they are referred to here as "participants."

Participants in the Importance and Testing Assessment Workshops included the Core Team members plus various experts selected for their particular expertise, their breadth of knowledge of the entire program, or their explicit knowledge of their organizations' role in the testing program described in the SCP. In addition, knowledge of the available site data and knowledge of the relationship of that data to performance-based parameters or related concerns was considered to be equally important and in many cases indispensable.

Twenty three participants were consulted for the Phase I application. Each individual was introduced to the concepts, methodology, and application of the Phase I approach and were requested to review the list of potential concerns to be evaluated. The individual participants were then to consult references, other participants, or any other source of information during the many assessment workshops. (A comprehensive listing of selected publications is included in the bibliography of Appendix B.)

Almost all of the 23 participants were involved in one or more of the six other assessment meetings of the TPT (see Appendix B). These six meetings (i.e., Performance Assessment Panel, Unsaturated Zone Hydrology Panel, Saturated Zone Hydrology Panel, Geochemistry Panel, Waste Package Panel, and the Gas Flow Panel) were held by the TPT to develop and evaluate influence diagrams, the assessment methodology, and the construction of the "simple" total-system-performance model that is planned to be used in the Phase II approach of the TPT. More than 40 experts participated in these workshops, in addition to the 23 participants directly involved in the Phase I effort. The meetings, meeting dates, and those in attendance at these workshops are tabulated in Appendix B.

In addition, several of the participants were active in related studies, such as the Calico Hills Risk/Benefit Analysis and Exploratory Shaft Alternatives Study. Effective assessment procedures and assessment results from those other projects were frequently considered during the Importance and Testing Workshops to supplement resources provided by the Core Team.

The twenty three workshop participants were drawn from the organizations participating in the Yucca Mountain Project (YMP) because the Core Team, the Technical Project Officers, and the Department of Energy considered that these personnel possessed the specific expertise needed and were available under the time constraints imposed on the Phase I approach. Although it was considered to be desirable to obtain "unbiased" external experts who have had little or no prior involvement with the YMP, it was recognized that the use of external experts was not practical in meeting the Phase I study schedule because of the amount of technical material that would need to be learned by the expert before being able to participate in the assessment process. An expert must be able to consider both the site specific and the regional information, and integrate this information while evaluating, refining, and responding to the assessment questions and process. It is planned in the Phase II application of this study to utilize a broader range of experts from both within and external to the Project, particularly on specific topics which are of great importance for evaluating the sites suitability.

Avoiding Biased Assessments

The existence of biases–conscious and unconscious—in subjective probability assessment is well known (Tversky and Kahneman, 1974), and there are generally accepted procedures for dealing with these biases (Spetzler and Staël von Holstein, 1975; McNamee and Celona, 1987).

For example, experts often assess a probability distribution with a range that is narrower than their actual level of uncertainty. Another common case is that experts anchor on a single number and assess some uncertainty around that central point, resulting in a distribution that does not reflect their true state of uncertainty. Such biases may be mitigated by focusing first on the extremes of a probability distribution, and then assessing its central values. These techniques were used whenever practical in the assessment process.

Another type of bias occurs when participants are motivated to provide judgments that differ from their conscious beliefs. For example, a salesperson who is asked to assess sales next year may give a low estimate if they are likely to be rewarded if their sales exceed the estimate. Such motivational biases may have existed in these workshops, considering that some participants have personal or organizational interests in particular tests. In general, the Core Team was convinced that even though this bias was possible, each participant made a sincere effort to provide honest judgments.

A third type of bias occurs when assessing low probabilities. It is difficult for many people to understand the difference between, for example, a probability of .001 and .0001. Nevertheless, many of the probabilities assessed by the group were for events with likelihoods in this range. Assessment aids were prepared to show the differences between such numbers. (See Fig. 2-2.) In some cases it was possible to break down low-probability events into conditional events. For example, the probability of a PC affecting waste isolation over the next 10,000 years might be very low. Rather than assessing the probability directly, the experts were asked to assess the probability at three levels:

- The probability that the PC is present today
- The probability that it will be present in the next 10,000 years, *given* that it is present today
- The probability that it will affect waste isolation, *given* that it will be present in the next 10,000 years.

For example, if each of these had probability .01, which is an assessment that is feasible for most people, then the joint probability of all three events occurring is .000001, which probably could not be assessed accurately by most people.

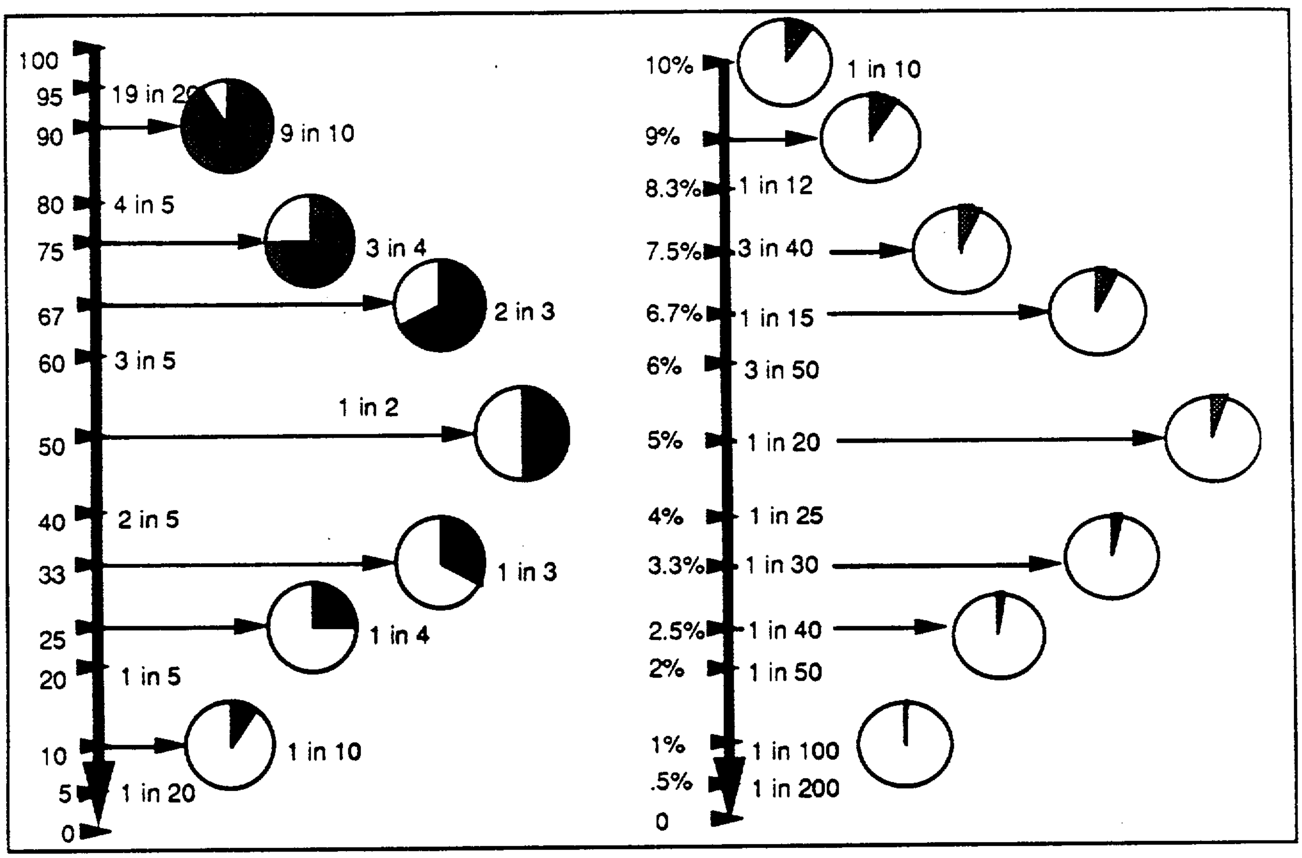

Figure 2-2a. Graphic aids for assessing "medium-probability" events. These graphic aids were used to portray the relative magnitude of probabilities in the range 0.5 to 100 percent. *Source: John Lathrop of Strategic Insights.*

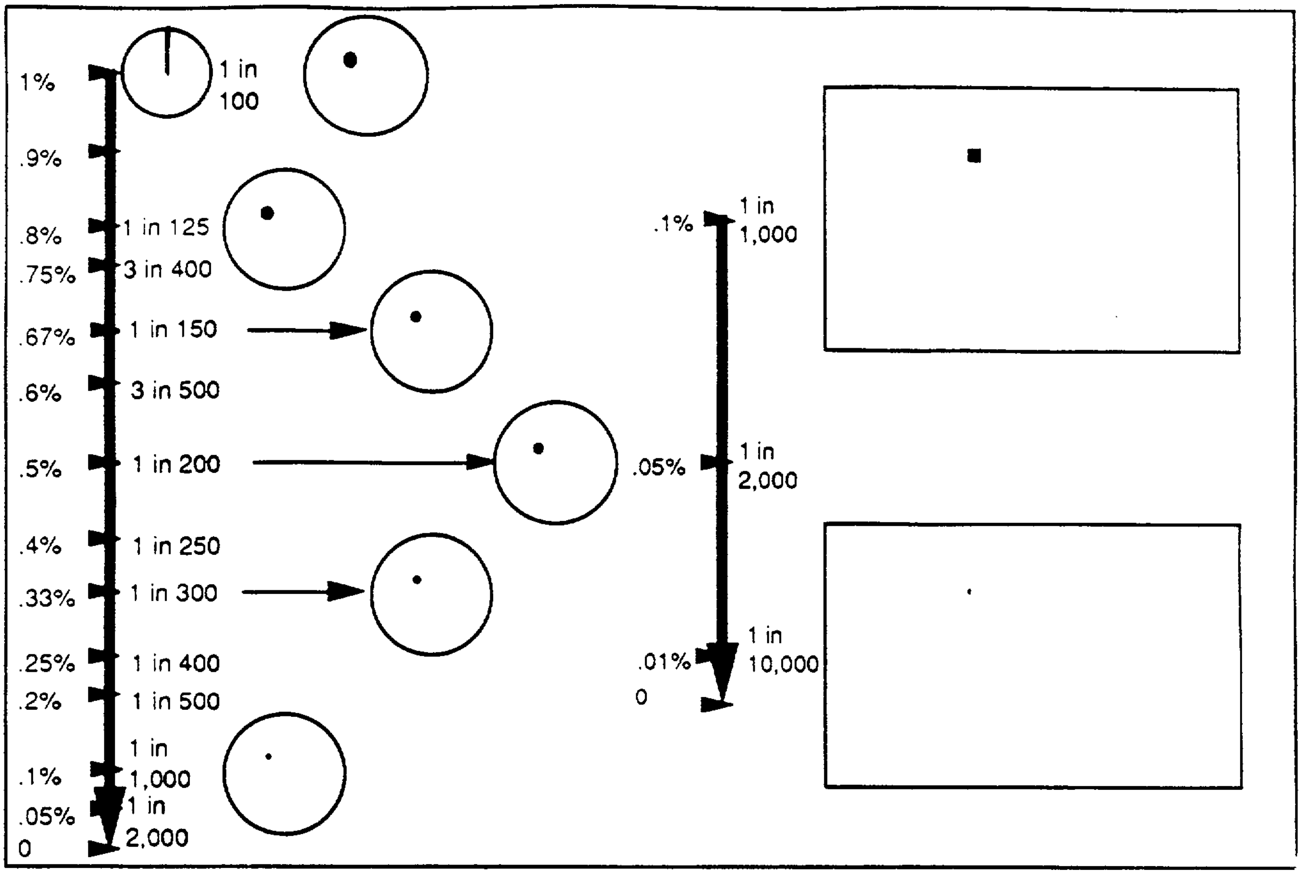

Figure 2-2b. Graphic aids for assessing "low-probability" events. These graphic aids were used to portray the relative magnitude of probabilities in the range 0.01 to 1 percent. Low probabilities are generally difficult to assess without a graphic or other type of assessment aid for reference. *Source: John Lathrop of Strategic Insights.*

Of course the assessment process takes time, especially when the process is carefully conducted to avoid bias. Generally it took at least an hour to assess a probability distribution—more when the process was new or when there was appreciable group discussion.

Balloting Process

The assessment process was the same for all potential concerns. First participants discussed the definition, measure, and assessment threshold suggested by the Core Team. This clarified any questions participants had about the PC and its measure. Often background information was shared, such as the reasons for including the potential concern on the list to be assessed.

Next the assessment questions were discussed and refined, if necessary. Then a printed "ballot" was distributed, with the assessment question clearly specified on the ballot. (A sample ballot is shown in Volume II, Appendix C.) For example:

Assessment question: What is the probability that expected pre-waste-emplacement ground-water travel time from the repository to the accessible environment is less than 1,000 years along the fastest path of likely and significant radionuclide transport?

Using their best judgment, the participants estimated an initial probability distribution for the assessment question as specified on the ballot. (Any participant who felt unqualified to make a particular assessment was instructed to abstain, which was noted on the ballot.) Most of the PCs involved an aggregation of many related issues, scenarios, and concerns, and the workshop participants were instructed to incorporate all relevant information and experience in making their assessments.

Next, probability assessments were recorded and projected on an overhead screen where all participants could observe the range of judgments. Then a discussion was led by the facilitator, information was exchanged, and rationales were given for various individual assessments, especially for the highest and lowest values. (See Appendix C for discussion.) A final ballot was then cast and collected, and the individual assessments were recorded to complete the assessment process.

Aggregating Individual Judgments

Because participants with diverse information about the PCs and possible tests were invited to the workshops, it was natural that they would differ about the probabilities of particular events. For example, Table 2-1 shows the diverse probabilities assessed for ground-water travel time.

Table 2-1. Sample individual probability assessments of expected ground-water travel time being less than 1000 years

Participant	Assessed probability
1	0.1
2	0.0001
3	0.01
4	0.05
5	0.0001
6	0.0001
7	0.001
8	0.0001
9	0.01
Average	0.019
Geometric mean	0.0015

The assessed probabilities in Table 2-1 range from .0001 to .1, spanning three orders of magnitude, or a factor of 1,000. After considering the appropriateness of both the arithmetic and the geometric means of the assessed values, the workshop participants concluded that the geometric mean (that is, the antilogarithm of the arithmetic mean of the logarithms of the individual probability values) constituted an appropriate summary statistic to represent the group assessment. For example, if there were three assessments of .1, .01, and .001, the arithmetic mean is .037 and the geometric mean is .01. It is important that the participants *chose* to use this averaging technique to represent their collective judgment. However, all judgments were retained, and sensitivity analyses were conducted to determine how much the testing priorities change if the analysis were based on an individual high or low judgment, rather than on geometric means.

Documenting the Proceedings

The person who recorded the workshops "took notes" in large-font text, with the video display projected onto an overhead screen, which was visible to all participants. Definitions, discussion points, probability assessments, and their rationales were recorded. This documentation was particularly useful in preparing written summaries of the meetings and for this Phase I report. (See Appendix B.) A spreadsheet was used to record all assessments, so that the geometric mean and other statistics could be computed immediately and made available to the participants during the discussions.

Lessons Learned

The assessment process evolved over the course of the workshops. There were several lessons learned that can increase the efficiency of subsequent assessment sessions:

- The preparation prior to the workshops by the Core Team or a small subset of workshop participants is greatly beneficial.
- Holding workshops off-site (away from interruptions) is also key to their success. Providing breaks for phone calls is necessary, but it is important to keep everyone participating except on breaks.
- The role of facilitator is critical. The facilitator must keep the pace and focus on the question at hand, summarize main points for the recorder, review them on the screen with workshop participants, and conduct assessments in an unbiased, professional manner.
- Using the computer models and projection equipment to show the implications of assessments is helpful. The models are handy tools for "on-line analysis" to show the implications of assessments and the sensitivity of results to differing opinions during the workshops.
- Creating a pleasant atmosphere for hard work, focused discussion, quality assessments, and building a team spirit also contributes significantly to meeting productivity. Everything from comfortable seating, U-shaped table arrangements, good lighting control for screen visibility, and after-hour socializing should be considered.

- Workshops take time! These activities require a substantial commitment of time and resources. Preparation, documentation, and "immediately-after-the-workshop" summary and analysis produce the greatest value for the effort invested in the workshop.

All things considered, the analytic method and assessment process worked well together. The next chapter summarizes the results of the assessment workshops and analysis results.

3. Assessment and Analysis Results

This chapter presents the details of how the five-step analytic method was implemented, the results of the assessment process, and the results of the analysis and Phase I prioritization. This chapter is organized in the order of the five analytic steps discussed in Chapter 2.

Phase I provided immediately usable results over a relatively short period of time. Some of those results might change as refinements are made during the planned Phase II effort. However, most of the insights that emerged during Phase I will remain valid.

Step 1: List Potential Concerns

Sources of Potential Concerns

The Core Team and workshop participants generated a set of 32 potential concerns (PCs) that were specific to the Yucca Mountain site for quantitative analysis in the Phase I study. These are listed in Table 3-1, along with definitions, reference numbers used to identify the PCs in Appendices A, C, and D, and short titles used to label the PC in tables and figures.

These 32 PCs were derived from several sources. Primary among these were the postclosure potentially adverse conditions (PACs) set forth in the siting guidelines of 10 CFR Part 60.122(c) and 10 CFR 960 and the disqualifying conditions in 10 CFR Part 960. Only postclosure issues from the regulations were included because waste-isolation performance was the measure of site suitability that was adopted as a prioritization criterion in this study. Because the focus of this study is on early tests that could detect potentially unsuitable conditions at the Yucca Mountain site, the favorable conditions from the siting guidelines (10 CFR Part 60) and the qualifying conditions (10 CFR Part 960) were not considered in this analysis.

In almost all cases the definitions of potentially adverse conditions differ slightly between 10 CFR Parts 60 and 960. For the most part, the Core Team chose the wording that resulted in the best definitions and associated measures for assessment purposes for each individual PC. It is intended that the definitions encompass the intent of the concerns as expressed in both 10 CFR Part 60 and 10 CFR Part 960.

The set of 32 PCs also incorporates a set of site-specific concerns that are not addressed in the regulations but were identified during an elicitation meeting held for this purpose in Las Vegas, Nevada, on February 8, 1990. Participants in that meeting were asked to consider their own concerns as well as those raised by the State of Nevada, the Edison Electric Institute, the Electric Power Research Institute, the Nuclear Regulatory Commission, and the general public (see the bibliography in

Table 3-1. List of potential concerns

Title of potential concern (PC)	PC number	Short title	Definition
Gas-flow radionuclide	1	Gas	Potential gas-phase release through the unsaturated zone to the accessible environment
Complex geology–Gaseous	2.2	CG-Gas	Incorrect prediction of gaseous-phase releases (by 10% of the baseline) because of expected or unexpected complexities in site geology or in the modeling of the site
Complex geology–Aqueous	2.1	CG-Aq	Incorrect prediction of aqueous-phase releases (by 10% of the baseline) because of expected or unexpected complexities in site geology or in the modeling of the site
Direct intrusion	H3	Intrusion	Direct intrusion of emplaced waste during drilling for water or economic resources
Expected GWTT<1000y	6	GWTT	Ground-water travel time (GWTT) less than 1,000 years along fastest path of likely and significant radionuclide transport
Oxidizing GW in host rock	4	Eh	Presence of chemically oxidizing ground water in repository host rock
Climate effect on Rn transp	5	Climate	Effects of possible future climate change on the site unsaturated-zone hydrologic system
Human intrusion effects-geohy.	H1	HI-Geo	Potential for future human activity to change hydrologic conditions at the site
Natural resources	H4	Nat res	Presence of potentially economically recoverable natural resources at the site
Perched water	8	Perched	Presence of perched water at or above the repository level
UO_2 Solubility	26	UO_2 sol	Solubility of UO_2 in ground water within the repository host rock
Past igneous act., site effects	16	Volcan	Potential effects of direct igneous intrusion into the repository
Reactive GW chem (EBS)	3	TDS	Presence of chemically reactive ground water in the repository host rock
Usable water in CA: SZ	H8	Use H2O	Potential for future ground-water withdrawal to affect the site saturated-zone hydrologic system
Water-table rise: 200m	9	200m rise	Potential for water-table rise at the site > 200 meters due to future climatic or tectonic change
Therm/rad effects: corr. steam	24	Steam	Generation of corrosive steam in the repository environment due to heat released by emplaced waste
Future mining	H6	Fut mine	Potential effect of future mining near the site on unsaturated-zone hydrologic system at the site
Therm/rad effects: resat. flux	23	Resat flux	Potential effects on waste isolation due to thermally induced drying and subsequent resaturation near the repository
Past active tectonism (faulting)	13	Faulting	Potential effects on waste isolation due to future faulting within the repository
Rock & GW complex engr.	19	Cmpl engr	Presence of rock or geohydrologic conditions at the site requiring complex engineering methods to construct and close the repository
Geomorphic processes, past eros.	25	Erosion	Occurrence of future erosion rates at the site sufficient to affect waste isolation

Table 3-1. List of potential concerns *(continued)*

Therm/rad effects: permeab. chg.	21	Perm chg	Potential for thermally induced permeability increases in the host rock near the repository
Tect eff. on reg. GW flow:UZ	11	Tect UZ	Potential effects of future tectonics on the site unsaturated-zone flow system
Past mining	H5	Old mine	Potential effects of past mining at the site on the unsaturated-zone flow system
Therm/rad effects: sorb. zeol.	22	Sorb Zeo	Potential thermally induced alteration of sorbing minerals near the repository
Tect eff. on reg. GW flow:SZ	12	Tect SZ	Potential effects of future tectonics on the site saturated-zone flow system
Water-table rise: 20m	10	20m rise	Potential for water-table rise at the site > 20 meters due to future climatic or tectonic change
Tectonic-induced lakes	15	Lakes	Possible occurrence of tectonic-induced lakes at the site
Past igneous act., CA effects	17	Volc CA	Potential effects of igneous activity near the site on waste isolation
Sorp/rock strength reduction	20	Rock str	Present rates of ongoing mineralogic change at the site
Rock cond. beyond RAT	18	Rock>RAT	Presence of rock conditions at the site requiring engineering methods beyond reasonably available technology
200m depth infeasible	7	200m depth	Inability to maintain a repository depth of at least 200 meters below land surface

Appendix B). Thus, the initial list of PCs was much longer, but some were eliminated because they did not apply to Yucca Mountain or because they did not affect waste isolation, and some were combined with others to avoid duplication or to simplify the assessment process. Details about the sources, references, and correlations of each PC can be found in Volume II, Appendix A.

Definitions, Measures, and Thresholds for the Assessed Potential Concerns
The disqualifying and potentially adverse conditions in the regulations generally are not quantitative. In order to facilitate quantitative assessment of the importance of the PCs, the Core Team defined each PC in terms of a quantitative measure that could serve as an indicator of site performance in the presence of the PC. Although these definitions and measures were sometimes revised during the assessment workshops, care was taken to ensure that an internally consistent set of PC definitions and measures was used throughout the analysis. The quantitative PC measures were intended to be representative surrogates for the PCs, as defined in this analysis, and are not intended in any way to be criteria for evaluating the presence of the PACs or disqualifying conditions as listed in the regulations.

For example, consider the regulatory guidance given for PC #3: Reactive Ground-water Chemistry (from 10 CFR Part 960-4.2.2) :

> 960.4-2-2 Geochemistry, (c) Potentially Adverse Conditions
> (1) Ground-water conditions in the host rock that could affect the solubility or the chemical reactivity of the engineered-barrier system to the extent that the expected repository performance could be compromised.

The definition used in this analysis is very similar to the PAC:

> **Definition:** Present ground-water conditions in the repository host rock that could affect the solubility or chemical reactivity of the engineered-barrier system.

Each PC was defined to capture the "flavor" of the corresponding regulation. Volume II, Appendix C provides a detailed discussion of the definitions of all PCs.

In order to translate the general definition of a PC into a form amenable to assessing the probability that the PC is present, the Core Team developed a surrogate measure of chemically reactive ground-water conditions and then determined a degree of chemical reactivity that could "compromise" repository performance. The following measure was one of two developed to reflect the concerns in 10 CFR Part 960-4.2.2. In this case the measure was developed with the assistance of workshop participants:

> **Measure:** Total dissolved solids (TDS) of ground water that could potentially contact the engineered-barrier system.

In some cases (e.g., ground-water travel time) the measure was specified in the regulations. In other cases (e.g., reactive ground-water chemistry) surrogate measures (e.g., TDS) were developed by the Core Team or in the assessment workshop. These quantitative measures provided the basis for subsequent probability and consequence assessments.

The presence or absence of potential concerns is usually a matter of degree. Thus, a specific value (magnitude) of the measure, denoted as the "assessment threshold," was established. "Present" was defined as conditions at the site being above or below the threshold, as appropriate. Conditions on the other side of the threshold were considered to indicate that the PC is not present. The assessment threshold was specified by the source of the concern (e.g., 10 CFR Parts 60 or 960 or both), by the Core Team, or by the workshop participants. Initial assignments of the thresholds were often revised by workshop participants.

In the case of reactive ground-water chemistry, the threshold was set equal to

> **Assessment threshold:** TDS = 10,000 parts per million.

TDS values exceeding 10,000 parts per million were regarded by the workshop participants as indicating the presence of a major concern about the ground-water chemistry at the site. Such a major concern could lead to high uncertainty in predicting aqueous-phase radionuclide releases, based upon possible SCP waste-package designs or the stability of the waste form.

The individual PC definitions, measures, and assessment thresholds are listed in Volume II, Appendix C. The measures and assessment thresholds listed there are not intended to quantify or otherwise revise or modify the regulations. They were devised only for the purposes of assessing test priorities. Furthermore, the assessment thresholds are in no way indicative of conditions that make repository performance unacceptable. They simply indicate points that would raise a concern that there *could* be a significant impact of the PC on postclosure waste isolation, although not necessarily enough impact to violate the EPA limit for cumulative curies released. The expected magnitude of the effect on cumulative curies released could be estimated quantitatively, for example, by a total-system-performance model.

The choice of assessment thresholds is discretionary. If one chooses a level of the threshold that is highly detrimental to waste isolation, then the probability that the PC exists (i.e., that the measure exceeds the assessment threshold) would be low. Conversely, if one chooses a threshold less detrimental to waste isolation, then its probability of being present would be higher. Ideally, one would choose thresholds such that they all had the same effect on waste isolation. This was not practical for the Phase I analysis but could be done with the aid of sensitivity studies conducted using a total-system-performance model, such as that planned to be used in Phase II study.

Initially, the Core Team attempted to set the assessment thresholds at a level that would raise questions about site suitability (i.e., a level at which it would be prudent to make detailed performance model calculations, to consider mitigation measures, or to reconsider the suitability of the site). However, it was difficult to establish such a level for most PCs, because the threshold levels were so far above the currently assessed site conditions. Therefore, the thresholds were set at lower levels where there *could* be significant effects on postclosure waste isolation. Although, the impacts of exceeding the threshold varied greatly among the PCs, the assessments performed in Step 2 quantified and accounted for that variation.

Finally, some conditions identified in 10 CFR Parts 60 and 960 are indicated by the regulations to be concerns whenever they are present; others are concerns only when they are present *and* have a "significant effect on waste isolation." The assessment definitions in this report use both types of definitions, although the *ultimate measure of either type of concern is their potential effect on waste isolation*. The assessment definitions were modified and refined, as necessary, to best match the extensive knowledge and technical experience of the participants in the assessment. This issue is discussed further in the description of Step 2.

Defining the Effect on Waste Isolation

When 10 CFR Part 60 and 10 CFR Part 960 refer to a condition as being "significant to waste isolation," it is not clear from the regulation how to define the PC. For example, assume that a hypothetical PAC reads "the presence of rock types that could significantly affect waste isolation." Further, assume that only "red rocks" have any detrimental effect on waste isolation. Then the PC could be defined for assessment as either "presence of red rocks" or "presence of a sufficient quantity of red rocks that cause the estimates of repository performance to exceed the EPA release limits." The distinction is important because the testing discussed in Steps 3 and 4 is directed toward "finding out whether the PC is present or not." The first definition of the PC—the presence of red rocks—is a physical definition, and the second is a performance definition. In 10 CFR Part 60, the statement of each PAC is preceded by the wording, "The following conditions are potentially adverse conditions if they are *characteristic* of the controlled area or if they *affect isolation* in the controlled area" [emphasis added]. This implies that *either* the physical or the performance definition could be used.

In general, physical definitions make it easier for physical scientists to assess whether the condition is present and to predict the accuracy of tests to detect the condition if it is present. This is because they are able to judge the likelihood of a particular test "finding" the specified PC, rather than judging the likelihood of "finding" the condition *and* estimating how it might affect performance by a specified amount. Quantifying the effects of a particular condition on performance requires a thorough working knowledge of total-system-performance models, their limitations, and their results, in addition to a firm grounding in the physical science issues being addressed.

The approach taken by the Core Team on various PCs differed, depending on the workshop participants' abilities to make physical or performance-related assessments. This will be illustrated in Step 2, below.

Step 2. Assess and Rank the "Importance" of PCs

Equation Used to Compute Importance

The ranking in Step 2 is based on assessed probabilities and consequence judgments for each PC. "Importance" in this context is measured by the expected incremental detriment to waste isolation caused by each PC. This measure is computed from several assessments:

- The probability of occurrence of each PC under present site conditions
- The probability of its presence in the future given that it is present today
- The probability that it will affect waste isolation if it is present in the future
- The incremental detriment on performance attributed to each PC that is present, if it exceeds the assessment threshold and affects waste isolation.

It is this last term, the incremental detriment on performance, that would be assessed more reliably with the aid of a total-system-performance model or other method of quantitative analysis.

"Incremental detriment" refers to the incremental curies released to the accessible environment over 10,000 years, relative to a baseline case in which none of the 32 potential concerns was assumed to be present. The incremental detriment is quantified using the ratio of curies released by each radionuclide to the release limits set by the Environmental Protection Agency in 40 CFR Part 191, Appendix A, Table 1. A ratio of 1.0 means that the expected cumulative curies released over

10,000 years following closure of the repository for a particular radionuclide exactly matches the number in Table 1. For example, for carbon-14 the EPA limit is 100 cumulative curies released per 1,000 metric tons from the repository to the accessible environment. Although there may well be other effects of a particular PC on the suitability of the site, this study uses cumulative curies released over 10,000 years after closure as a proxy for the detrimental effects of the PC on waste isolation.

The limits set forth in EPA Table 1 were chosen to restrict releases to levels that would result in no more than 1,000 excess cancer deaths over 10,000 years from the disposal of 100,000 metric tons of reactor fuel (40 CFR Part 191.13). Because the potential repository is planned to accommodate roughly 70,000 metric tons of reactor fuel (DOE, 1988), the value 700 excess cancer deaths over 10,000 years was considered in this Phase I analysis to be equivalent to a ratio of 1.0 times the EPA release limits.

Using these basic concepts, the following variables were defined and assessed for each PC:

- A — The probability that the assessment threshold is exceeded and, therefore, that the PC is present today or was present in the past (usually taken to be the Quaternary Period)
- B1 — The probability that the PC will occur in the next 10,000 years, given that it is (or was) present
- B2 — The probability that the PC affects waste isolation, given that it will occur in the next 10,000 years
- C — The multiplier on baseline curies released, given that the PC affects waste isolation
- D — Baseline performance defined as the normalized curies released in the *absence* of *all* PCs.

Using these variables, importance is computed as follows:

Importance = A x B1 x B2 x (C-1) x D.

The term (C-1) appears rather than C alone because importance is defined in terms of incremental curies released, i.e., the curies in excess of the baseline releases, D.

Alternately, it may be useful to think of the importance of a PC as the product of two terms:

Importance = Probability that the PC is present x
Expected incremental consequences given that it is present.

Using the variable names from above:

P(PC present) = A

Expected consequences | PC present =
Expected increase in curies | PC present = B1 x B2 x (C-1) x D

where P() is defined as the probability of an event and | is read "given that."

This definition of importance provides a measure of the relative significance of each PC, as well as a measure of the expected value of resolving whether or not the PC is present. If one assumes that the consequences of the PC could be prevented after correctly discovering that it is present (either by mitigation or by abandoning the site), then the importance value equals the "expected value of perfect information" about the presence of the PC. Here "perfect" means 100-percent test accuracy, and "value" is measured in terms of potential curies avoided (Holloway, 1979). Also, no specific measures to prevent the release of radionuclides were evaluated. Thus, the terms "mitigate" or "abandon" are intended only to represent the range of possible actions that might be taken in response to a detection.

Relationship Between Assessed Probabilities and the Concerns Expressed in the Regulations

As defined above, A, the probability that the PC is present, is the probability that the assessment threshold is exceeded. When making the assessments for Step 2, the Core Team tried to maintain consistency between the assessment threshold and the concern as expressed in the regulations. If the regulation said "past presence of a condition," then the probability A was defined as the probability that the PC occurred in the past. If the regulation mentioned "future presence," then the probability A was defined as the probability that the PC would be present in the future. In the latter case, the probability B1 was not needed, and B1 was set equal to 1.0.

This flexibility was acceptable in the assessment of importance in Step 2. However, when assessing test accuracy in Step 4, the definition of probability A had to be consistent with and related to the conditions being investigated in the tests. Therefore, some definitions of the probabilities A, B1, and B2 had to be changed from Step 2 to Step 4. (These will be noted later, in Step 4.) This was done in a way to preserve, to the extent possible, the assessments made in Step 2.

Assessed Importance Parameters

Table 3-2 lists the assessments made during the two Importance Workshops. Appendix B in Volume II provides the dates, locations, and participants in the workshops. The numbers listed are the "group" assessments, i.e., the geometric mean of responses given by all individuals for each assessed variable. The individual judgments for each potential concern are presented in Appendix C, along with the PC definitions, quantitative measures, assessment thresholds, and exact assessment questions for each variable. Appendix C also summarizes salient points made in the discussion that took place during the assessment workshops.

Table 3-2. Assessed importance parameters

Potential concern (PC)	PC number	Prob. that threshold exceeded (PC pres.)	Prob. PC pres. next 10000 yr | PC pres.	Prob. PC affects WI | future PC	Multiplier on perform. | PC affects WI	Baseline perform. without PC
		A	B1	B2	C	D
Gas-flow radionuclide	1.1	6.2e-01*	1.0e+00	1.0e+00	5.3e+04	1.7e-06
Complex geology–Gaseous	2.2	3.2e-01	1.0e+00	1.0e+00	4.2e+04	1.7e-06
Complex geology–Aqueous	2.1	2.9e-02	1.0e+00	1.0e+00	1.4e+05	1.7e-06
Direct intrusion	H3	2.6e-02	1.0e+00	2.2e-02	3.1e+04	1.7e-06
Expected GWTT<1000y	6	1.5e-03	9.7e-01	5.8e-01	9.9e+02	1.7e-06
Oxidizing GW in host rock	4	8.6e-01	9.9e-01	1.6e-01	3.1e+00	1.7e-06
Climate effect on Rn transp	5	2.2e-03	1.0e+00	3.2e-01	1.5e+01	1.7e-06
Human intrusion effects-geohy.	H1	1.5e-03	1.0e+00	5.7e-01	1.1e+01	1.7e-06
Natural resources	H4	2.2e-03	5.6e-03	6.5e-01	9.8e+02	1.7e-06
Perched water	8	2.6e-02	6.2e-01	1.2e-01	3.6e+00	1.7e-06
UO_2 Solubility	26	4.7e-03	9.7e-01	4.0e-01	2.7e+00	1.7e-06
Past igneous act., site effects	16	9.9e-01	2.2e-05	1.7e-01	4.3e+02	1.7e-06
Reactive GW chem (EBS)	3	3.6e-04	1.0e+00	2.1e-01	1.9e+01	1.7e-06
Usable water in CA: SZ	H8	9.5e-01	5.1e-02	9.0e-02	1.2e+00	1.7e-06
Water-table rise: 200m	9	1.3e-04	1.0e-02	8.0e-01	2.8e+02	1.7e-06
Therm/rad effects: corr. steam	24	1.2e-01	7.5e-01	1.0e-02	1.2e+00	1.7e-06
Future mining	H6	1.0e-05	1.0e+00	5.1e-01	2.1e+01	1.7e-06
Therm/rad effects: resat. flux	23	2.5e-02	3.9e-01	1.9e-02	1.2e+00	1.7e-06
Past active tectonism (faulting)	13	1.5e-01	1.6e-01	1.7e-03	1.4e+00	1.7e-06
Rock & GW complex engr.	19	1.3e-03	9.9e-01	1.5e-02	1.3e+00	1.7e-06
Geomorphic processes, past eros.	25	4.6e-05	6.4e-01	7.6e-02	3.1e+00	1.7e-06
Therm/rad effects: permeab. chg.	21	2.3e-02	8.6e-01	5.6e-03	1.0e+00	1.7e-06
Tect eff. on reg. GW flow:UZ	11	2.2e-04	1.7e-03	2.8e-01	1.7e+01	1.7e-06
Past mining	H5	8.4e-04	1.0e-04	6.5e-01	1.6e+01	1.7e-06
Therm/rad effects: sorb. zeol.	22	9.3e-03	6.9e-01	4.3e-03	1.0e+00	1.7e-06
Tect eff. on reg. GW flow:SZ	12	6.4e-03	6.0e-03	1.7e-02	1.8e+00	1.7e-06
Water-table rise: 20m	10	3.0e-02	2.0e-02	1.3e-02	1.0e+00	1.7e-06
Tectonic-induced lakes	15	6.0e-03	2.2e-04	5.8e-02	1.4e+00	1.7e-06
Past igneous act., CA effects	17	9.9e-01	2.2e-04	4.6e-04	1.2e+00	1.7e-06
Sorp/rock strength reduction	20	6.0e-05	1.0e+00	7.7e-04	1.1e+00	1.7e-06
Rock cond. beyond RAT	18	1.3e-04	1.0e+00	2.8e-04	1.1e+00	1.7e-06
200m depth infeasible	7	8.6e-04	1.0e+00	1.3e-04	1.0e+00	1.7e-06

*The entry 6.2e-01 is read, "6.2 times ten to the power of negative one: 6.2×10^{-1}

The assessment in Col. A is the probability that the PC is present now, in the past, or in the future, as specified in the definition of the PC. Column B1 lists the assessed conditional probability that the PC will be present at some time during the next

10,000 years, given that it is present today. This probability equals 1.0 when the event in Col. A is defined as "occurrence at some time during the next 10,000 years."

Column B2 lists the assessed conditional probability that the PC will affect waste isolation, given that it occurs at some time during the next 10,000 years. This probability equals 1.0 when the event in Col. A is defined as "present at a level that affects waste isolation." The numerical range of variation in this column spans four orders of magnitude.

Column C tabulates the estimated incremental increase in normalized curies released if the PC were present (i.e., the ratio of actual curies released to the EPA release limits). This was assessed as a multiplier on the baseline performance (Col. D). However, because the magnitude of the potential increases in releases for any radionuclide is uncertain, the number in Col. C was assessed as the expected (or mean) value of the multiplier, defined as the probability-weighted average of all possible effects due to the presence of the PC. For example, if the value in Col. C were 1.0, then there would be no increase in curies released due to the presence of the PC. A value of 1.1 means a 10 percent increase in cumulative curies released; a value of 10 means that cumulative curies increase nine-fold. The numerical range of variation in this column span more than five orders of magnitude.

Column D lists the baseline performance, which is defined to be the geometric mean of the set of each individuals assessment of expected cumulative curies released to the accessible environment over 10,000 years, normalized to the EPA release limit. The following assumptions were made when assessing this baseline level: (1) none of the PCs are present (i.e., all PC measures are below their respective assessment thresholds), (2) 10 percent of the emplaced waste packages fail in 10,000 years, and (3) all radionuclide transport to the accessible environment is by ground-water flow.

A value of 1.0 in Col. D means that the average releases to the accessible environment of all radionuclides would equal the cumulative release limits specified by EPA. Based on the conversion factor in 40 CFR Part 191.13, the assessed value of 1.7×10^{-6}, or 0.0000017 times the EPA release limits corresponds to a level of 0.0012 excess cancer deaths over 10,000 years (which is equivalent to one expected excess cancer death every 8 million years).

The entries in Cols. C and D were some of the most difficult assessments for the workshop participants to judge because of the need to consider so many interrelated factors when estimating performance. A total-system-performance model is planned to be used in Phase II of this analysis to compute estimates of releases. In fact, discussion during the workshops often revealed that individuals were providing estimates of the results they presumed would be produced by a total-system-performance model that used unbiased probability distributions as inputs. Here "unbiased" means that the probability distributions encompass the true state of current uncertainty about these input parameters and that there are no "conservative" assumptions in the model or input parameters. Also, the Phase II total-system-performance model, presumably could be capable of incorporating viable alternative conceptual models and statistical correlations among physical parameters.

Computed Importance of Potential Concerns

Table 3-3 summarizes the results of the Importance Assessment. Three columns of numbers are listed:

- Probability that the potential concern is present (A)
- Expected consequences of the PC if it is present (E)
- Computed importance of the potential concern (I).

The importance of a PC (I) is the product of the probability of occurrence (A) and its expected consequences (E).

The concerns in Table 3-3 are sorted from top to bottom in order of decreasing assessed importance. The most important PC, using this method is "Gas-flow Radionuclide," which is, specifically, the transport of gas-phase radionuclides (especially carbon-14) through the unsaturated zone. A close second is "Complex Geology," which is site geology of sufficient geohydrologic complexity as to cause a significant modeling underestimation of potential gaseous or aqueous releases. If a normalized release value of 1 corresponds to 700 excess cancer deaths in 10,000 years, then perfect information about these highest-ranked PCs could prevent between 5 and 39 expected excess cancer deaths over 10,000 years, assuming that no mitigation is possible unless the PCs are detected.

Figure 3-1 provides a graphic interpretation of the results in Table 3-3. The vertical bars graph the expected consequences *if each PC is present*. Thus, the vertical height of the bars indicates the relative magnitude of each PC, irrespective of the probability that it is actually present. The height is measured in terms of the expected increase in normalized curies released, *if the presence of the PC were confirmed*. The dots connected with a solid line indicate the importance of each PC. (The line is provided as a visual aid and does not imply a functional relationship.) Importance is interpreted as the *expected* consequences of each PC, taking into account the probability that the PC is present. The units are the same as for consequences: expected increase in curies released due to each PC (normalized to the EPA limits).

This chart illustrates the major factors that determine the importance of each PC. The first factor is the relative magnitude of the potential consequences of each PC if it is present. Because the importance of a PC is its probability of occurring times the expected consequences if it does occur, the relationship between the height of the dot and the height of the bar indicates whether the PC is likely to occur or not. For example, a dot near the top of a bar indicates that the PC is likely to be present; in this case the importance is determined by the relative magnitude of the consequence. On the other hand, if the dot is well below the top of a bar, then the

Table 3-3. Computed importance of potential concerns

Potential concern (PC)	PC num-ber	Prob. that PC is present	Expected consequence given PC is present	Importance to waste isolation
		A	E B1xB2x(C-1)xD	I AxB1xB2x(C-1)xD
Gas-flow radionuclide	1.1	6.2e-01	8.8e-02	5.5e-02
Complex geology–Gaseous	2.2	3.2e-01	7.0e-02	2.2e-02
Complex geology–Aqueous	2.1	2.9e-02	2.4e-01	6.9e-03
Direct intrusion	H3	2.6e-02	1.2e-03	3.0e-05
Expected GWTT<1000y	6	1.5e-03	9.3e-04	1.4e-06
Oxidizing GW in host rock	4	8.6e-01	5.5e-07	4.7e-07
Climate effect on Rn transp	5	2.2e-03	7.3e-06	1.6e-08
Human intrusion effects-geohy.	H1	1.5e-03	9.7e-06	1.5e-08
Natural resources	H4	2.2e-03	6.0e-06	1.3e-08
Perched water	8	2.6e-02	3.2e-07	8.4e-09
UO_2 Solubility	26	4.7e-03	1.1e-06	5.1e-09
Past igneous act., site effects	16	9.9e-01	2.7e-09	2.6e-09
Reactive GW chem (EBS)	3	3.6e-04	6.4e-06	2.3e-09
Usable water in CA: SZ	H8	9.5e-01	1.7e-09	1.7e-09
Water-table rise: 200m	9	1.3e-04	3.7e-06	4.8e-10
Therm/rad effects: corr. steam	24	1.2e-01	2.6e-09	3.2e-10
Future mining	H6	1.0e-05	1.7e-05	1.7e-10
Therm/rad effects: resat. flux	23	2.5e-02	2.0e-09	5.0e-11
Past active tectonism (faulting)	13	1.5e-01	1.7e-10	2.4e-11
Rock & GW complex engr.	19	1.3e-03	8.5e-09	1.1e-11
Geomorphic processes, past eros.	25	4.6e-05	1.7e-07	7.9e-12
Therm/rad effects: permeab. chg.	21	2.3e-02	2.2e-10	5.2e-12
Tect eff. on reg. GW flow:UZ	11	2.2e-04	1.2e-08	2.7e-12
Past mining	H5	8.4e-04	1.7e-09	1.4e-12
Therm/rad effects: sorb. zeol.	22	9.3e-03	1.4e-10	1.3e-12
Tect eff. on reg. GW flow:SZ	12	6.4e-03	1.3e-10	8.6e-13
Water-table rise: 20m	10	3.0e-02	9.5e-12	2.8e-13
Tectonic-induced lakes	15	6.0e-03	7.4e-12	4.5e-14
Past igneous act., CA effects	17	9.9e-01	4.2e-14	4.1e-14
Sorp/rock strength reduction	20	6.0e-05	1.8e-10	1.1e-14
Rock cond. beyond RAT	18	1.3e-04	5.9e-11	7.7e-15
200m depth infeasible	7	8.6e-04	2.2e-12	1.9e-15

PC is unlikely to be present. In fact, the probability of occurrence of the PC is the ratio of the height of the dot to the height of the bar.

The range in importance spans nearly 14 orders of magnitude, reflecting the group assessment that some PCs are much less important than others. Twenty seven of

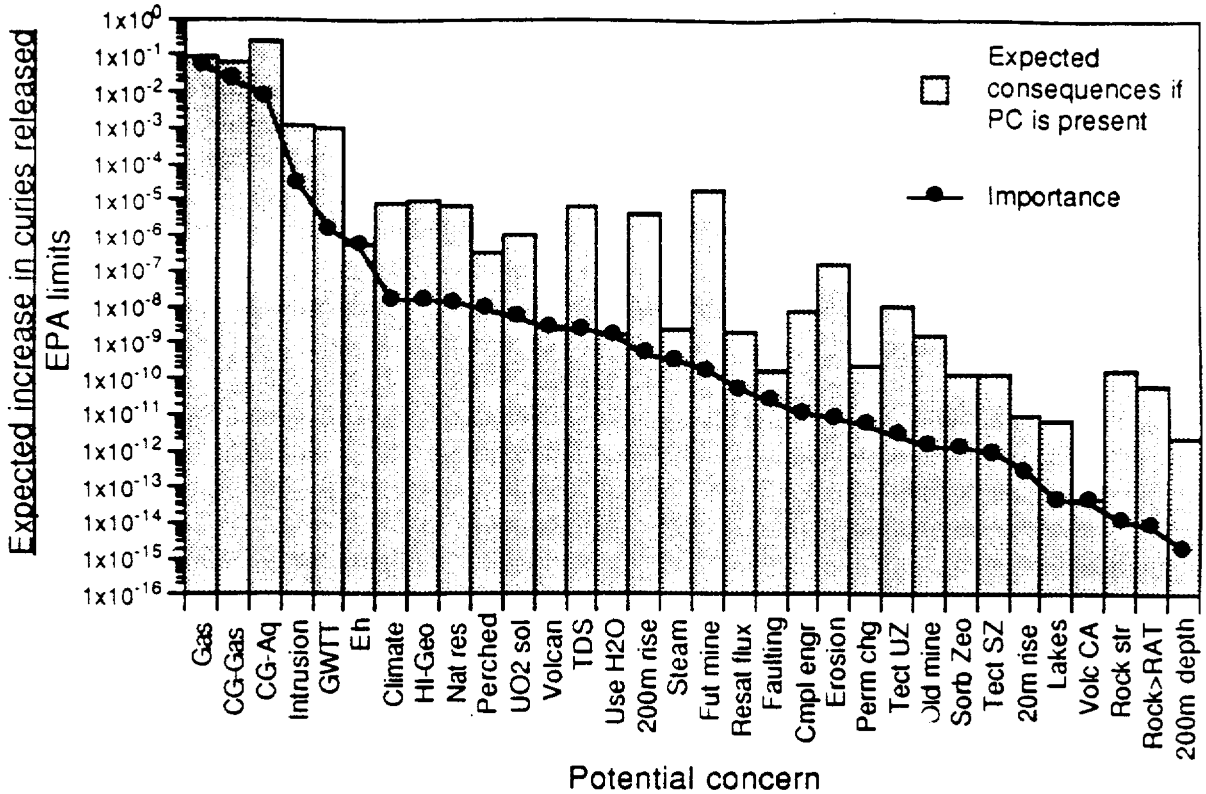

Figure 3-1. Importance and expected consequences of potential concerns. Vertical bars represent consequences *if* the potential concern (PC) is present, i.e., the increase in expected releases, normalized to the EPA limits. The dots and solid line represents the importance of the PC: the product of the probability that the PC is present and the expected consequences if it is present. According to 40 CFR Part 191.13, value of one on the vertical axis (top of the graph) would correspond to 700 excess cancer deaths over 10,000 years, or one excess death expected every 15 years (for a repository containing 70,000 metric tons of heavy metal). Abbreviations for PCs are found in Table 3-1; detailed descriptions are in Appendix C.

the 32 PCs have expected consequences below 10^{-3} consequence units, which is less than 0.7 excess cancer deaths in 10,000 years. The highest importance corresponds to 39 excess cancer deaths over 10,000 years. Therefore, the absolute importance of resolving uncertainty about any PC is low in terms of protecting public health and safety, given the judgments of the panel of experts about the probabilities and magnitude of releases associated with individual PCs.

Maximum Benefits of Testing

The importance line in Fig. 3-1 quantifies the maximum benefits of testing that are associated with discovering each potentially unsuitable site condition. The importance numbers represent the *maximum* number of incremental curies that could be avoided if the PC were detected. This assumes:

- The curies could and would be completely avoided (either by mitigation or by abandoning the site) if the presence of the PC were established.

- The test for the PC is perfect, that is, it has no chance of missing the PC if it is present and no chance of erroneously concluding that it is present if it is not.

No specific mitigation or abandonment strategies were evaluated explicitly; these are meant only to represent the range of possible responses to the detection of a PC. Also, it is important to recognize that abandonment cannot avoid the curies altogether because some risk is associated with current storage in pools at reactor sites, and it is likely that some risks would remain if an alternative storage site or method were selected.

Tests could be prioritized on the basis of the importance results (the dots) in Fig. 3-1. However, importance, as defined here, ignores the accuracy and false-alarm costs of the proposed tests. Both of those factors will be introduced in Step 4. However, the factors that contribute to the importance numbers in Step 2 strongly influence the subsequent priorities that emerge from completion of Steps 4 and 5. The final priorities derived in Steps 4 and 5, which consider test accuracies, are not substantially different from the results in Fig. 3-1. The only major differences occur when the costs (false-alarm or dollar costs) of the tests exceed their benefits.

The results in Fig. 3-1 do not account for potential consequences associated with the simultaneous occurrence of two or more PCs, which could result in synergistic processes that increase releases over the sum of the effects of individual PCs. Also, as stated earlier, the importance determined here is only in the context of early testing for unsuitable site conditions. The overall importance of testing would include several other considerations, as discussed in Chapter 1.

Sensitivity Analysis

The sensitivity of the results to the input judgments is important in any analysis. Because this analysis is based almost exclusively on expert judgment, analysis of the sensitivity of results to differences in judgments of individual workshop participants is an essential step.

The range of expert opinions on the probabilities, consequences, and test accuracies is of primary interest. The geometric mean of the set of individual judgments is used to represent the group consensus about inputs to the analysis. However, this single statistic does not show the effects of the range of opinion on importance assessments.

Figure 3-2 shows the range of the individual assessments of importance for each PC. The dots represents the geometric means of all individual responses, as shown in Fig. 3-1. The vertical bars define the range of the assessments by the workshop participants and are plotted as the highest and lowest individual assessments.

Figure 3-2 indicates that there was a wide diversity of opinion among group members about the absolute and relative importance of individual PCs. The

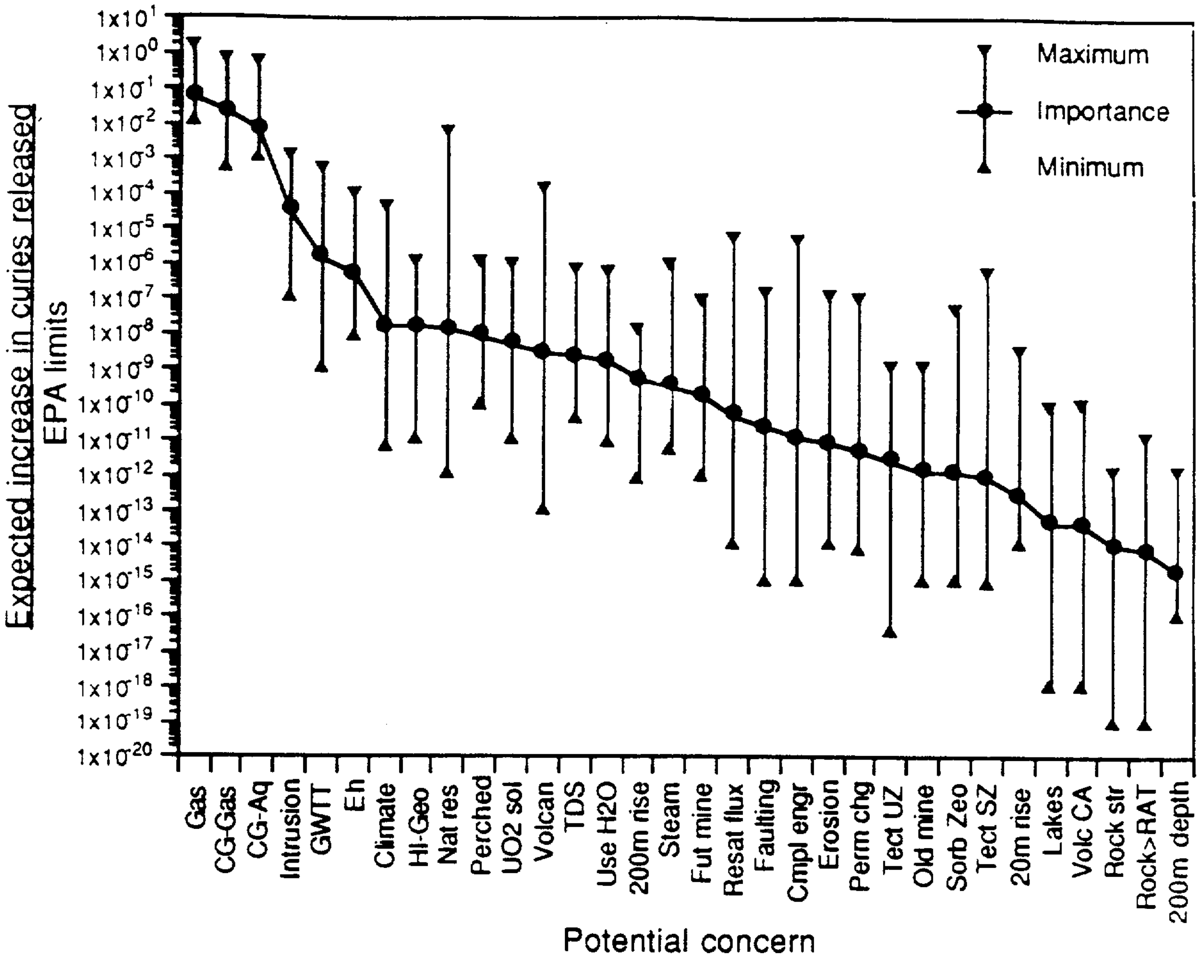

Figure 3-2. Sensitivity of importance to individual assessments. This graph shows the maximum and minimum importance numbers assessed by the participants in the assessment workshops. The value used to plot the solid line for importance is the geometric mean of the assessments of all participants, reproduced from Fig. 3-1. Based on this graph, essentially all workshop participants agree that Gas-flow Radionuclide and Complex Geology (Gaseous and Aqueous Pathways) are the most important potential concerns.

minimum and maximum importance judgments for single PCs range up to nine orders of magnitude. Contributing to this overall variation are the three orders of magnitude variation in the assessment of baseline curies released (Col. D in Table 3-2). As expected, various members of the group provided judgments that would modify the priorities established using geometric means. Generally, the group members agreed that the PCs "Gas-flow Radionuclide" through "Complex Geology–Aqueous" are the three most important. Furthermore, none of the participants gave any of the lowest-ranking 18 PCs ("200m water-table rise" through "200m depth infeasible") an importance number greater than they gave for any of the first 14 PCs.

Screening Important PCs
Based on their assessed importance, the PCs were grouped as follows: the six most important PCs designated, "Gas-flow radionuclide" through "Oxidizing ground water in host rock," had the highest potential for being associated with significant benefits from testing. "Climate effects on radionuclide transport" through "Usable water in the controlled area," are all about of equal importance, whereas, the importance of the remaining PCs decreases at a somewhat faster rate than the second group of PCs. The first two groups of 14 PCs were carried forward to the test accuracy assessment.

The Testing Workshop participants further screened the list of 14 PCs and the following PCs were either eliminated or combined with others for the reasons given:

- *Complex Geology–Gaseous (2.2).* Testing associated with this PC was judged to be adequately addressed by the Gas-flow Radionuclide PC.

- *Direct Intrusion (H3).* This PC, inadvertent drilling into a waste canister, was judged to be closely associated with the Natural Resources PC. The expected consequences of this PC were added to those for Natural Resources.

- *Human Intrusion Effects on Geohydrology (H1).* Workshop participants agreed that this issue is not amenable to the gathering of technical site data toward resolution; this is an issue for (if anything) socioeconomic analysis and development of a position paper on the topic.

- *Usable Water in the Controlled Area (H8).* Like PC #H1, this PC was eliminated because, in the view of workshop participants, there is little need for new testing or additional data. There currently is enough information (and enough data) for analysis, drawing correlations, and preparing a final position.

This left ten PCs that were assessed for test accuracy.

Step 3. Identify Tests that Address High-ranking PCs

Sources of Tests
The "tests" that address each PC are based on studies and activities identified and described in the Site Characterization Plan (SCP). The Core Team used the PARATRAC data base (SAIC, January 21, 1991) to provide an initial list of SCP tests for each PC. From the PARATRAC lists and a reference copy of the SCP, workshop participants created the lists of tests for each PC that are provided in Appendix D. Participants were not limited to suggesting the tests described in the SCP. Other tests were acceptable candidates if they were germane to investigating the PC.

In most cases, only a subset of the tests listed in the SCP or PARATRAC is appropriate for early testing to identify potentially unsuitable site conditions. Therefore, the lists of tests in Appendix D should be viewed as those tests that could be performed early to assist in early evaluation of site suitability.

Testing Packages
As mentioned in Chapter 2, most individual tests were grouped into "suites" or "packages" of tests that would likely be conducted together. Sometimes there was only a single test package, in which case the calculations of test benefits in Step 4 represented an upper bound on the information that could be obtained from testing. In other cases two or three levels of test packages were evaluated, for example:

- Level 1. No new boreholes; use existing data
- Level 2. New surface-based boreholes plus existing data
- Level 3. New surface-based boreholes and existing data plus the Exploratory Shaft Facility.

Generally, these packages were constructed so that both the testing accuracy and costs increased progressively from one level to the next. Thus, evaluation of these packages can show the marginal benefits of increasing test accuracy with more extensive testing. No attempt was made by the workshop participants to prioritize tests within a test package or a testing level, nor were the test packages ordered with respect to possible scheduling constraints. The tests were prioritized on the basis of the information they could provide and the accuracy with which they were judged to be able to provide the information. The tests within each package or level are listed in Appendix D for each of the PCs for which test accuracy was assessed.

Step 4. Assess the Accuracy, Benefit, and Cost of Tests

This step assesses the accuracy of each test package and identifies the most effective package for each PC. Step 4 also provides the essential information for the comparisons and rankings in Step 5.

This section of the report begins with a discussion of how test accuracy is defined and assessed. Next, the results of the assessment are discussed, beginning with the assessed test accuracies and concluding with overall measures of the detection benefit and false-alarm cost of each test package. The detailed assessments are listed in Appendix D in Volume II.

Quantifying Test Accuracy
As mentioned in Chapter 2, the accuracy with which each test detects the presence of a potential concern is quantified using two numbers:

- The conditional probability of "finding" the PC, given that it is present
- The conditional probability of "finding" the PC, given that it is not present.

These are referred to as probabilities of true positive and false positive (or false alarm), respectively. They are both important because both can lead managers to take actions that affect waste isolation, either correctly or incorrectly. Effective management considers both the "up-side" (i.e., the possibility of a true positive) and "down-side" (i.e., the possibility of false positive) before conducting tests.

Most tests do not directly detect the presence of a PC. Rather, they generate data that must be analyzed and interpreted before a conclusion can be drawn about the presence of a PC. Therefore, the event "finding a PC" was defined to mean that the final position report, written after all data gathering and analysis activities are completed, either concludes that the "PC is present" or it is not.

In some cases, the true-positive and false-positive probabilities were assessed directly for each package of tests, using a probability tree as shown in Fig. 3-3. The first event represented in the tree is whether the PC is present or not. In most cases this probability (P1) was derived from the Importance Workshops; in general, P1 is the same as A in Table 3-2. The probability P2 in Fig. 3-3 is the probability of finding the PC given that it is present (true positive). Probability P3 is the probability of (incorrectly) finding the PC given that it is not present (false positive or "false alarm"). In other cases, however, the assessment was broken down to lower-level probabilities, as is discussed in the following paragraphs.

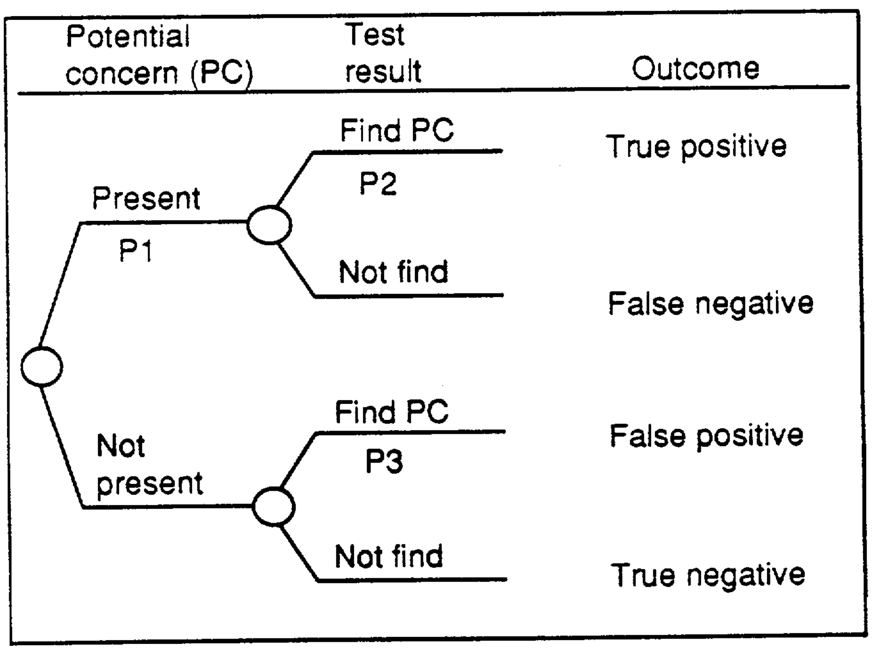

Figure 3-3. Probability tree showing test accuracy assessments. This analysis quantifies accuracy using two probabilities: the conditional probability that a test will correctly conclude that a potential concern is present, when, in fact, it *is present* (P2), and the conditional probability that the same test will erroneously conclude that the PC is present when, in fact, it *is not present* (P3). These probabilities, P2 and P3, are referred to as the probabilities of true and false positive, respectively. In cases where the PC is clearly present or not, P2 and P3 were assessed directly by participants in the Testing Workshops. In other cases, when the PC was measured by a continuous variable, probabilities P2 and P3 were computed from assessments of testing accuracy over the range of possible values for the continuous variable. The *a priori* probability that the PC is present, P1, was assessed in the Importance Workshop.

Assessing Test Accuracy for Continuous Variables
The assessments illustrated in Fig. 3-3 are best suited to PCs that are *discrete*, that are, either present or not (e.g., the presence of a new magma body beneath the region). However, in many cases the PC is measured by a *continuous* variable, such as total dissolved solids, oxidation potential (Eh), or expected ground-water travel time along a specified path. In this case the presence of the PC is determined by whether the continuous measure for the PC is "above" or "below" the assessment threshold. In such cases, one can assume that the test "reports" a value for the continuous measure, which is then interpreted and a conclusion is drawn regarding the presence or absence of the PC. In these instances, test accuracy was quantified using a more detailed procedure from which the probabilities P2 and P3 were computed.

Consider the example of PC #4, Oxidizing Ground Water in the Host Rock, for which the measure is oxidation potential expressed as Eh in millivolts (mV). The first step in the assessment was to quantify the current degree of uncertainty associated with the measure under present site conditions (the "prior" probability distribution, where "prior" in this context refers to probabilities of occurrence before testing). In most cases, five points on the cumulative probability distribution were assessed, corresponding to 1%, 10%, 50%, 90%, and 99% cumulative probabilities. A smooth curve was then fitted to these points, as shown in Fig. 3-4.

Next, the continuous cumulative distribution was approximated using a three- or four-step discrete approximation (i.e., the stair step function shown in Fig. 3-4). (See Holloway, 1979, p. 215.) A discrete approximation is made to simplify the assessments of test accuracy and computations of probabilities P2 and P3 in Fig. 3-4. The three-step approximation in Fig. 3-4 is interpreted as follows: current uncertainty in oxidation potential is adequately characterized by three values: Low Eh (200 mV), Medium Eh (400 mV), and High Eh (600 mV). The probabilities assigned to these three values are .27, .35, and .38 respectively.

Test accuracy is then assessed conditioned on each of these three values. This conditional assessment accounts for the possibility that test accuracy may vary as a function of the value of the variable being tested.

Because Eh is a continuous variable (although it is approximated using three discrete values here), the possible results of testing are also assumed to be a continuous function. This is illustrated in Fig. 3-5, which shows a hypothetical lognormal probability density function for the results of testing a variable whose true value is T. The figure illustrates a case where there is 80-percent confidence that, if the true value is T, the test will produce a value somewhere between $T \div F$ and $T \times F$. For example, assume the true oxidation potential (Eh) is 200 mV. Under this assumption, Eh is no longer uncertain; the assumption is that it is exactly equal to 200 mV. Further, assume that the uncertainty in test results is appropriately represented by a lognormal probability distribution and that testing in this case is

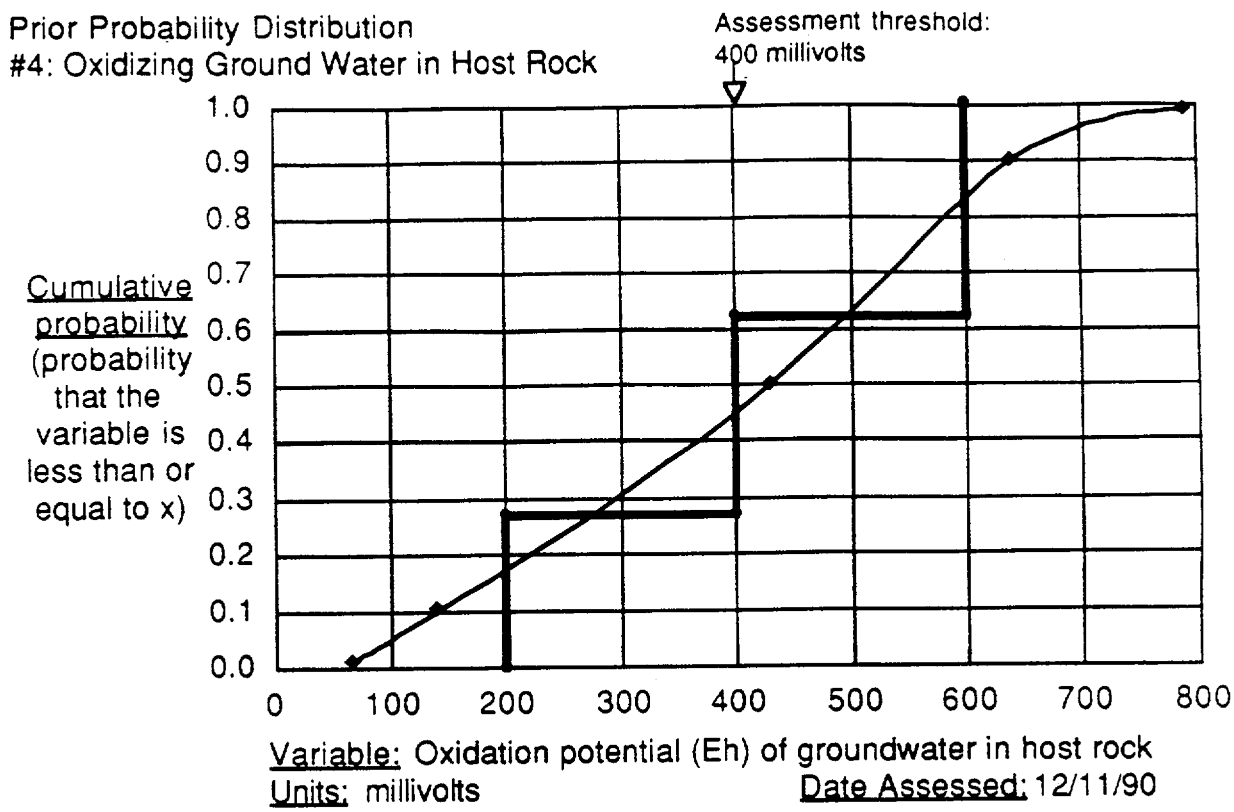

Figure 3-4. Cumulative probability distribution showing current uncertainty regarding oxidation potential. In cases where a potential concern was measured by a continuous variable, such as oxidation potential (Eh), the Testing Workshops began with an assessment of the current level of uncertainty associated with the variable. Five points were generally assessed, corresponding to 1, 10, 50, 90, and 99 percent points on a cumulative probability distribution. These are shown as the points on the curve in the figure. A smooth, continuous curve was then drawn through the assessed points. This curve was then approximated using the 3-point stair-step function shown in the graph. This established three representative values of the variable (200, 400, and 600 mV in this case), with associated probabilities of .27, .35, and .38, respectively. The accuracy of testing was assessed at each of these three levels of Eh. At least one of the points on the assessed curve and its stair-step approximation is on either side of the assessment threshold, which is 400 mV in this case. This assures that test accuracy will be assessed for conditions when the potential concern is present (i.e., when Eh exceeds the assessment threshold) and when it is not present (i.e., Eh is less than 400 mV). The assessment thresholds were established in the Importance Workshops.

"accurate to within a factor of 2.0." This means that if the true value is 200 mV, then there is an 80 percent chance that the test will indicate a value between 100 and 400 mV (200 ÷ 2.0 and 200 x 2.0 mV, respectively). In this case the factor F in Fig. 3-5 is 2.0. This factor, called the "F factor," provides a convenient way to specify the endpoints of the 80-percent confidence interval for testing accuracy.

This method of quantifying test accuracy assumes that the departure of the test results from the true value, i.e., the uncertainty shown in Fig. 3-5, can be represented either by lognormal probability distribution, that is, that logarithm of the test result can be represented by a normal (Gaussian) probability distribution.

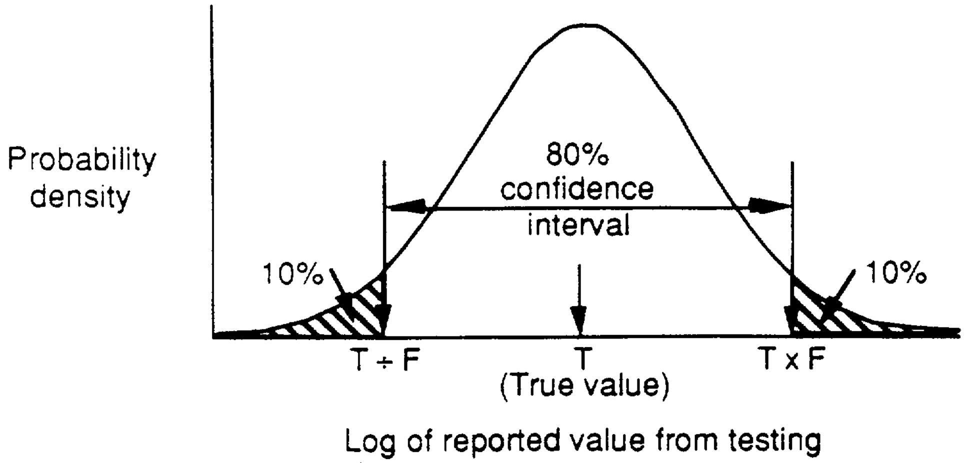

Figure 3-5. Representation of test accuracy for a lognormally distributed continuous variable. The figure depicts the range of results—reported values—from an imperfect test, given that the true value of the variable is T. In the case illustrated, there is an 80-percent chance that the reported value from the test will fall in the interval between T÷F and TxF. The "F factor" characterizes the accuracy of the test, in cases where the uncertainty in test results can be represented using a lognormal probability distribution. For example, assume that the principal investigator for a test for oxidation potential judges the test to be "accurate to within a factor of 2.0." If the uncertainty is lognormally distributed and the true value for oxidation potential is 200 mV, then there is an 80-percent chance that this particular test will report a value between 100 and 400 mV. In cases where the uncertainty in testing is normally distributed, test accuracy was assessed using "plus or minus" ranges rather than F factors.

Among other things, this means that the probability distribution is symmetric (in fact, it is "bell" shaped) when the horizontal axis is the logarithm of the test variable.

In many cases, the workshop participants believed that the density function in Fig. 3-5 was not lognormally distributed or it was not symmetric, even in "log space." For example, two participants said that if the true value of Eh were 200 mV, then there was an 80 percent chance that the value reported from testing would be between 100 and 300 mV. In other words, they had high (80 percent) confidence that the test would report a value within ±100 mV of the true value. In this case, the uncertainty in test results was assumed to be normally distributed. Participants had the opportunity to express their judgment on test accuracy using either this "plus or minus" value, using an "F factor," or by specifying *any* values for the low and high ends of the 80-percent confidence interval.

In the Testing Workshop, conditional confidence intervals were assessed for all three values of oxidation potential: 200, 400, and 600 mV. For this particular example, the probability distribution is assumed to be normal, rather than lognormal. The group consensus about the three 80-percent confidence intervals are ±100, ±133, and ±134 mV, respectively. Three corresponding probability density functions are shown in Fig. 3-6, with the 400 mV assessment threshold added to each. The distributions become increasingly "wider" from top to bottom in the figure as the confidence intervals expand. As a percentage of the true value,

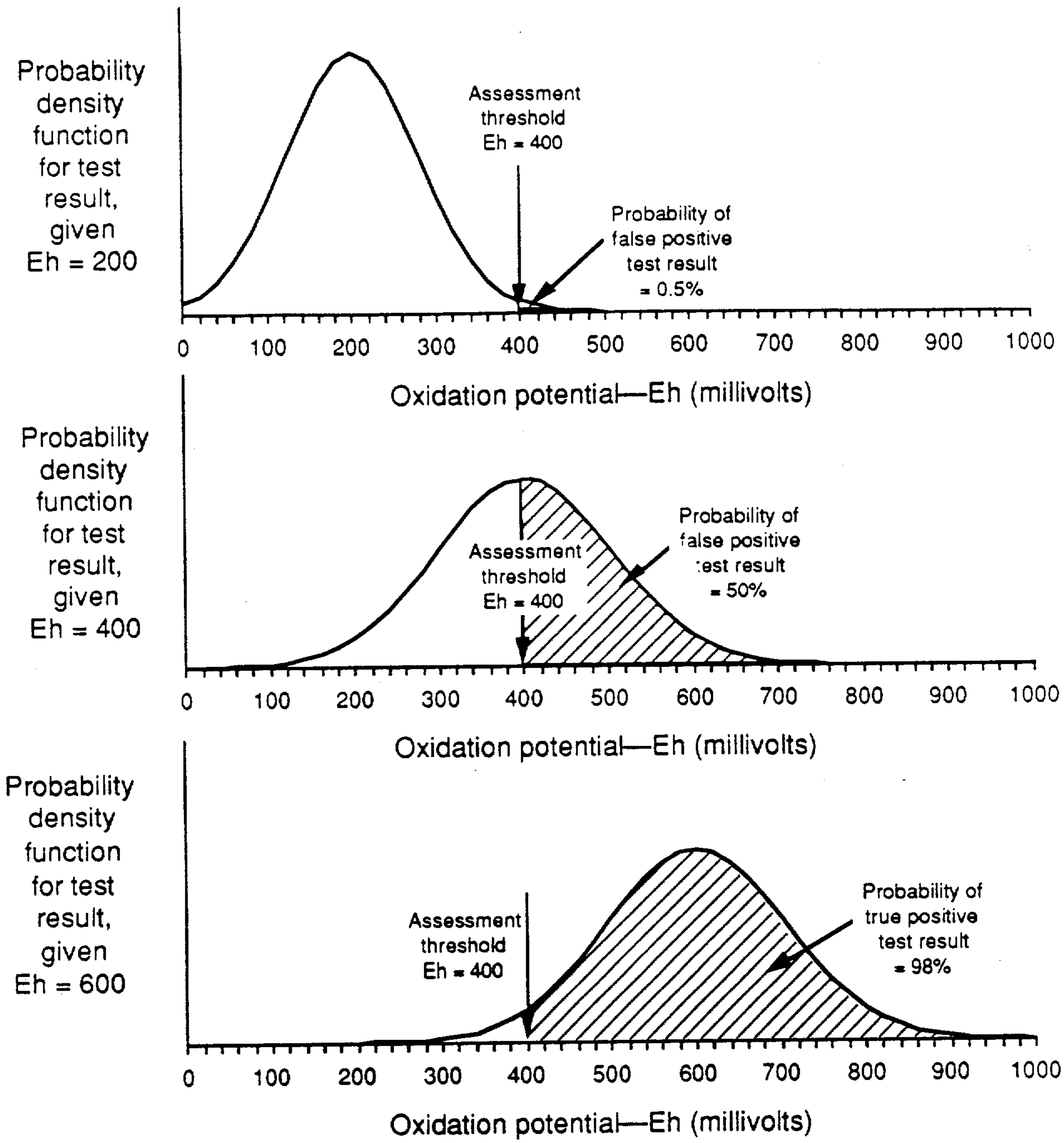

Figure 3-6. Conditional probabilities of exceeding the assessment threshold for low, medium, and high values of oxidation potential. These probability density functions illustrate the assessed uncertainty in test results for three possible true values of Eh: 200, 400, and 600 mV. If the true value is 400 mV (middle graph), then the test is equally likely to report a result above or below 400 mV. Further, there is an 80-percent probability that the report will be between 267 and 533 mV. The shaded area in each graph is the conditional probability that the test's principal investigator will report a value above the assessment threshold of 400 mV for oxidation potential. Such a report in the top case, when the true Eh is less than 400 mV is considered a "false positive" indication that the potential concern is present. The report in the bottom graph that Eh is greater than 400 mV is a "true positive" indication of the presence of the concern, because Eh in the lower graph is 600 mV. The designation of the shaded area in the middle graph as a "false" positive is arbitrary. If this were considered a true positive, then lower values would result for the conditional probabilities of both true and false positives. Graphs like these were used to compute the conditional probabilities of true and false positive (P2 and P3 in Fig. 3-3) for concerns measured by continuous variables.

however, the confidence intervals contract as the true value increases from 200 mV to 600 mV.

There is a shaded area in each drawing, which represents the conditional probability that the test indicates an Eh in excess of 400 mV, given the true value. If the true value is below 400 mV (top figure), then a reported value exceeding 400 mV is a "false positive." If the true value exceeds 400 mV (bottom figure), then a reported value in excess of 400 mV is a "true positive." The middle case, where the true value of Eh equals 400 mV, the shaded area is defined arbitrarily as a false positive and is grouped with the top case.

Finally, the shaded areas in the top two illustrations in Fig. 3-6 are combined to give the conditional probability of a false positive, P3 = 29 percent. As noted in the discussion of Fig. 3-4, the probabilities associated with Eh values of 200, 400, and 600 mV are .27, .35, and .38, respectively. The probability of false alarm, P3, is using rounded numbers; (.27 x .005 + .35 x .50) / (.27 x .35) = .28. Without rounding, P3 = .29 or 29 percent. The shaded area in the bottom illustration is the conditional probability of true positive, P2 = 98 percent.

These results are dependent on how one treats the middle case in Fig. 3-6. If the shaded area in the middle graph were considered to be a true positive, rather than a false positive, the results would be P2 = 74 percent and P3 = .5 percent, rather than the 98 and 29 percent, respectively. One might attempt to avoid this ambiguity regarding how to treat the 400 mV threshold by setting the threshold equal to 399 mV rather than 400 mV. Then the 98 and 29-percent probabilities for P2 and P3 would obtain. However, if the threshold were 401 mV then the 74 and .5-percent probabilities would apply. (The magnitude of the sensitivity of P2 and P3 to the assessment threshold would decrease if a greater number of representative values—more stair-steps—were used to approximate the cumulative distribution function in Fig. 3-4.)

This apparent dilemma highlights an important feature of characterizing the accuracy of testing. The accuracy probabilities, P2 and P3, are coupled, and they are intimately related to the assessment threshold. Changing the assessment threshold (or how the middle case in Fig. 3-6 is handled in this particular example) changes P2 and P3, and both probabilities change in the same direction. Furthermore, one can force P2 to take on any value from 0.0 to 1.0, simply by changing the threshold values. Of course, if the probability of a true positive, P2, is set very high, then the associated false-alarm probability, P3, also will be very high.

In this analysis the assumption is made that "action" to avoid the potential release of curies is taken at the level of the assessment threshold. The optimal choice of a decision point (e.g., the lowest Eh for which the action is taken) depends on P1, P2, P3, and the relative consequences of true and false positive test results. Such calculations are beyond the scope of this report. However, the numbers assessed in the workshops could be used to make this type of decision.

In summary, the assessments of test accuracy for PCs that are measured by continuous variables were more complicated than those for PCs that are measured only by their presence or absence. The assessment for continuous variables began by assessing a prior probability distribution for the measure of the PC (e.g., Eh for oxidation potential) and then choosing a discrete "stair step" that approximated the continuous distribution. This was followed by an assessment of the accuracy of testing using F factors or "plus-or-minus" ranges for each of the discrete approximations. Finally, the probabilities of true and false positive results (P2 and P3, respectively) were computed as illustrated in Fig. 3-6. The Core Team concluded that this process provided a more thorough and accurate quantification of test accuracy for continuous variables than attempting to assess P2 and P3 directly.

Accuracy Assessment Results

Table 3-4 lists the assessed and computed test accuracies, along with two other numbers used to compute the benefits of testing. The left-hand column lists the potential concerns carried forward from Step 2, Table 3-3. The participants in the Testing Workshops identified multiple levels of testing for the following PCs:

Complex geology
- CG-AQ-BH: Level 1: Borehole studies
- CG-AQ-ESF: Level 2: Borehole studies plus Exploratory Shaft Facility (ESF)

Expected ground-water travel time (GWTT)
- GWTT: Level 1: No drilling and using existing core samples
- GWTT-BH: Level 2: New drilling and coring plus Level 1 data
- GWTT-ESF: Level 3: ESF plus Level 2 data

Perched water
- Perched-BH: Level 1: New drilling
- Perched-ESF: Level 2: ESF plus Level 1 data

Past igneous activity–volcanism
- Volcan-Rate: Level 1: Studies of the rate of volcanic events
- Volcan-Magma: Level 2. Investigation of the possible presence of a new magma body beneath the Yucca Mountain region.

Each of these is treated as a separate package of tests and is listed as a separate row of Table 3-4. See Appendix D in Volume II for more details.

The first column of numbers in Table 3-4 is the prior probability (i.e., before testing) that the PC is present and is, for the most part, duplicated from Col. A in Table 3-2. Numbers in italics indicate changes from Table 3-2 and include the following:

- Cases where new information was assessed in the Testing Workshops leading to new prior probability distributions that superseded the prior probabilities assessed during the Importance Workshops (complex geology, climate, and oxidizing ground water).

Table 3-4. Inputs to the calculation of test benefits

Potential concern (PC)	PC number	Short name	Prob. that threshold is exceeded (PC present)	Consequences given PC present	Prob. of finding PC given PC pres.	Prob. of finding PC given not pres.
			A*	E*	F	G
Gas-flow radionuclide	1.1	Gas	*8.1e-1*	*6.7e-2*	94%	23%
Complex geology–Aqueous	2.1A	CG-Aq-ESF	*3.7e-2*	2.4e-1	86%	5%
Complex geology–Aqueous	2.1	CG-Aq-BH	*3.7e-2*	2.4e-1	66%	14%
Natural res. & dir. intrusion	H4+H3	Nat res	2.2e-3	*1.2e-3*	56%	1%
Climate effect on Rn transp	5	Climate	*2.5e-1*	7.3e-6	56%	6%
Expected GWTT<1000y	6B	GWTT-ESF	1.5e-3	9.3e-4	82%	1%
Expected GWTT<1000y	6A	GWTT-BH	1.5e-3	9.3e-4	77%	4%
Expected GWTT<1000y	6	GWTT	1.5e-3	9.3e-4	66%	12%
Oxidizing GW in host rock	4	Eh	*3.8e-1*	5.5e-7	98%	29%
Perched water	8A	Perched-ESF	2.6e-2	3.2e-7	63%	2%
Perched water	8	Perched-BH	2.6e-2	3.2e-7	60%	6%
UO_2 Solubility	26	UO_2 sol	4.7e-3	1.1e-6	50%	3%
Past igneous act., site effects	16	Volcan-Rate	*2.2e-5*	*1.2e-4*	58%	0%
Reactive GW chem (EBS)	3	TDS	3.6e-4	6.4e-6	62%	0%
Past igneous act., site effects	16A	Volc-Magma	4.4e-5	1.3e-7	62%	29%

** Italics indicate revised assessments*

- Cases where the definitions used to assess the variables A, B1, B2, C, or D were changed in the Testing Workshops, and the revised values were recorded in different columns in the Testing Workshop version of Table 3-2 (gas flow, natural resources, and volcanism–rate).

The first type of change affected the importance number computed in Table 3-3, but the order of PCs in the table (sorted in order of decreasing importance) did not change. The second type of change did not affect the importance calculation and did not affect the importance ordering.

Column E in Table 3-4, the expected consequences given that the PC is present, is derived from Table 3-3. It is repeated here for convenience because it is used in calculations in a subsequent table. Again, italicized numbers highlight changes from Table 3-2. These occurred for two reasons:

- Two potential concerns were combined (natural resources and direct intrusion).

- The definitions for variables A, B1, B2, C, or D changed in the Testing Workshops, and the values were recorded in different columns (gas flow and volcanism–rate).

The two right-hand columns give the assessed or computed test accuracies: the conditional probabilities of true and false positive results. The "best" tests are those having high values in Col. F and low values in Col. G.

Figure 3-7 plots the assessed and computed test-accuracy results. One point is plotted for each PC and each level of testing, as listed in Table 3-4. The vertical axis is the conditional probability of detecting the PC when, in fact, it is present (from Table 3-4, Col. F), and the horizontal axis plots the probability of erroneously detecting the PC even though it is not present (from Table 3-4, Col. G). Consistent with the objectives listed in the previous paragraph, the best tests appear in the upper left hand corner of the figure.

For example, consider the three points representing three testing levels for GWTT. The shift to the upper left corner as the level of testing increases, reflects the assessments by the workshop participants that there is a noticeable improvement in test accuracy as one progressed from currently available data, to additional borehole studies, to conducting the ESF tests.

The discussion following Fig. 3-6, explained how the probabilities of true and false positive test results, P2 and P3, are coupled, and that they depend on the level of the assessment threshold. For example, if one changed the threshold for GWTT, the three GWTT points would shift toward the upper right or lower left corner. However, they would all shift together, and their separation relative to the upper left corner would remain roughly the same.

Some insights can be drawn from Fig. 3-7. For example, one can compare the relative accuracy (proximity to the upper left corner) for any combination of tests. The GWTT tests provide a good example of how accuracy improvements can be quantified and compared. The two volcanism test packages (i.e., tests related to the rate of volcanic events and tests to determine the presence of a new magma body) provide a dramatic example of variations in test accuracy in Fig. 3-7 (see points Volcan-Rate and Volc-Magma, respectively). However, Fig. 3-7 should not be the only basis for setting test priorities because it leaves out two very important considerations:

- The probability that each PC is present
- The relative consequences of correctly or erroneously detecting PCs.

These considerations are introduced in the following paragraphs.

Calculated Benefits of Testing

The benefits and costs of the tests can be divided into two types: those that depend on the outcome of the test (e.g., "finding" or "not finding" an unsuitable condition) and those that accrue just because the test is conducted and do not depend on its outcome (e.g., its dollar cost or the credibility it could add to a license application). Only the first type is discussed in this report, because the primary emphasis of this report is on early tests that could detect potentially unsuitable site conditions. The analytic method and spreadsheet models could accommodate benefits and costs that

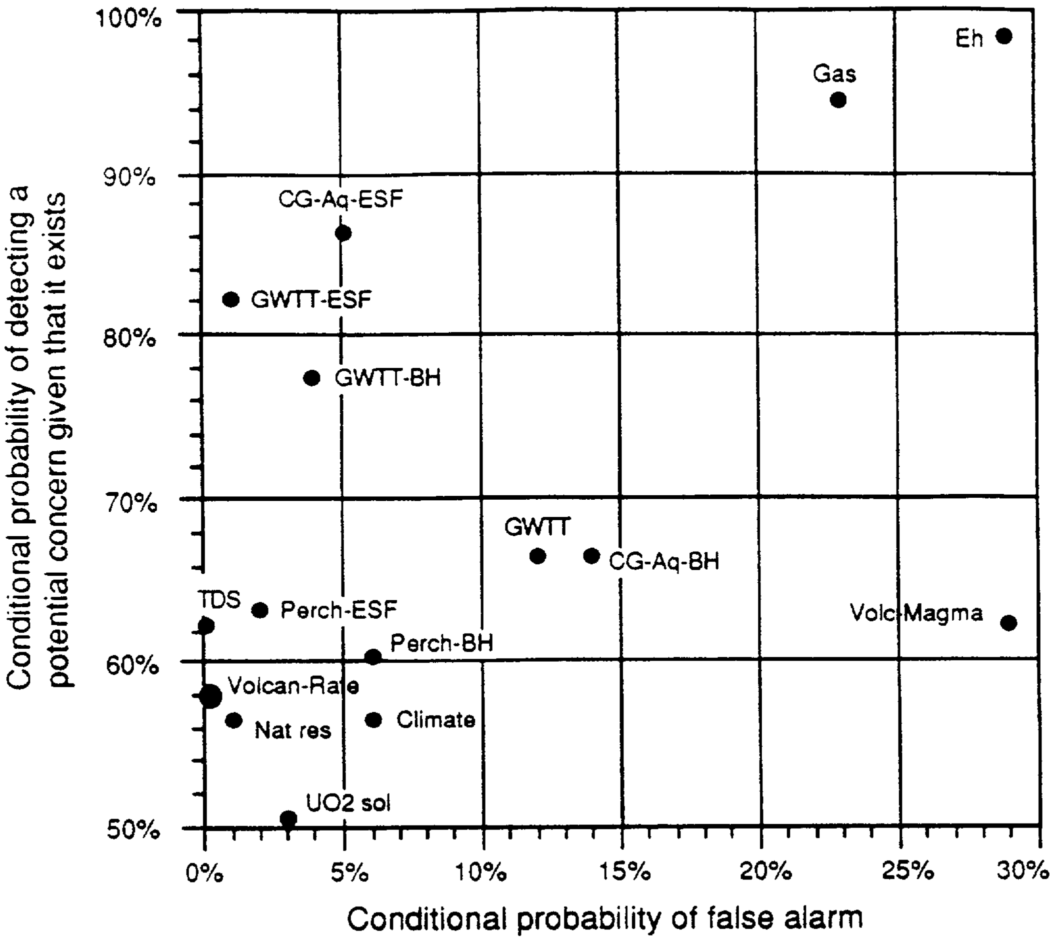

Figure 3-7. Conditional probabilities of detection and false alarm for proposed tests for important potential concerns. These conditional probabilities were assessed or computed for each test and are listed in Table 3-4. The closer the test is to the upper left corner, the greater overall accuracy it has. In general, workshop participants attributed greater accuracy to tests in the Exploratory Shaft Facility than to borehole (BH) tests. See for example the improvement obtained by the Ground-water Travel Time (GWTT)-ESF test compared to GWTT-BH. While illustrative of test accuracy assessments, this graph does *not* provide a good basis for prioritizing tests because it does not take into account the probability of occurrence and consequences for each potential concern.

do not depend on the outcome, but they were not assessed during the Phase I analysis.

The benefits of testing for potentially unsuitable site conditions are quantified using the results of the importance assessments from Step 2. Detection benefits are measured as the maximum detriment, defined here as the postclosure cumulative curies released over 10,000 years that could be *avoided* by accurately detecting the presence of a PC. The actions taken to avoid this detriment were not evaluated; only the maximum possible "benefit" was used in the calculations. This may overstate the benefits of some tests if no mitigating actions are possible; however, it is always possible to abandon the Yucca Mountain site and avoid the detriment of the PC altogether. (Abandoning the site, of course, would involve a trade-off among the benefits and detriments of continued on-site storage of accumulated waste and of siting a repository elsewhere, none of which is considered here.)

Table 3-5 lists four columns of numbers that take into account all of the factors for determining test benefits (within the scope of this analysis). Column I is the importance column from Table 3-3, using the updated numbers listed in Table 3-4. As indicated earlier, Col. I is computed as follows:

Importance = Probability that the PC is present x
Expected incremental consequences given that it is present.

Column J lists the expected detection benefits of correctly detecting the PC if it is present. This number is the maximum possible benefit of conducting an imperfect test to detect potentially unsuitable site conditions. The detection benefit, expressed in units of potential curies avoided, is computed as:

Detection benefit of test =
Probability that the PC is present (A) x
Probability that the test finds the PC, given that it is present (F) x
Expected curies potentially avoided given that the PC is present (E).

This equation assumes that, if the test finds the PC, action (mitigation measures or abandoning the site) will be taken to avoid the consequences in terms of curies released. The detection benefits in Col. J range over ten orders of magnitude.

Column K represents the possible "down-side" of testing, specifically the potential consequences of a false alarm. The "false-alarm cost" is assumed to be *proportional to* the detriment caused by the PC, which means that the false-alarm cost is proportional to the PC's "importance" and "detection benefit." This proportionality assumption was made by the Core Team because the action that could be taken for a highly detrimental PC is likely to be more costly or more dramatic than the action taken for a relatively minor PC. To account for this dependence on the detriment caused by the PC, false-alarm costs were assigned in proportion to the incremental detriment computed in Step 2. (While the Core Team assumed proportionality, management might make a different assumption.)

The assumption is made that the same action is taken whether the PC is "found" correctly or incorrectly. Further, it is assumed here that the consequences of this incorrect action are proportional to the curies avoided. For example, the cost of needless mitigation measures would be larger if the potential releases are large than if they are small. Based on this assumption, the calculation of Col. K is:

Table 3-5. Calculated benefits of testing

Potential concern (PC)	PC number	Short name	Importance to waste isolation(WI)	Detection benefit of test	False-alarm cost	Net benefit of test
			I	J	K	L*
			AxE	AxExF	(1-A)xExG	
Gas-flow radionuclide	1.1	Gas	5.5e-2	5.1e-2	2.9e-3	4.6e-2
Complex geology–Aqueous	2.1A	CG-Aq-ESF	8.8e-3	7.6e-3	1.2e-2	5.8e-3
Complex geology–Aqueous	2.1	CG-Aq-BH	8.8e-3	5.8e-3	3.2e-2	2.4e-3
Natural res. & dir. intrusion	H4+H3	Nat res	2.5e-6	1.4e-6	1.1e-5	3.1e-7
Climate effect on Rn transp	5	Climate	1.8e-6	1.0e-6	3.4e-7	8.9e-7
Expected GWTT<1000y	6B	GWTT-ESF	1.4e-6	1.1e-6	7.6e-6	3.5e-7
Expected GWTT<1000y	6A	GWTT-BH	1.4e-6	1.1e-6	3.5e-5	-2.2e-6
Expected GWTT<1000y	6	GWTT	1.4e-6	9.2e-7	1.1e-4	-9.2e-6
Oxidizing GW in host rock	4	Eh	2.1e-7	2.0e-7	9.8e-8	1.8e-7
Perched water	8A	Perched-ESF	8.4e-9	5.3e-9	7.8e-9	4.1e-9
Perched water	8	Perched-BH	8.4e-9	5.0e-9	1.8e-8	2.9e-9
UO_2 Solubility	26	UO_2 sol	5.1e-9	2.6e-9	3.7e-8	-1.0e-9
Past igneous act., site effects	16	Volcan-Rate	2.6e-9	1.5e-9	4.0e-11	1.4e-9
Reactive GW chem (EBS)	3	TDS	2.3e-9	1.4e-9	4.8e-16	1.3e-9
Past igneous act., site effects	16A	Volc-Magma	5.8e-12	3.6e-12	3.8e-8	-3.5e-9
* L = (Wt1xJ–Wt2xK) / (Wt1+Wt2)			Relative weights	10 Wt1	1 Wt2	

False-alarm cost =

Probability that the PC is not present (1–A) x
Probability that the test finds the PC given that the PC is not present (G) x
Expected curies avoided if the PC is detected (E).

As with the detection benefits in Col. J, the false-alarm costs range over many orders of magnitude. In most cases, high detection benefits are correlated with high false-alarm costs, because of the assumption that false-alarm costs are proportional to the expected curies avoided if the PC is found.

One other factor is included in the calculation of the net benefits of testing: the relative value of the detection benefits compared to the false-alarm costs. Because these benefits and costs are assumed to be *proportional* to the expected curies avoided (but not equal to them), proportionality constants, or weights, are required to combine benefits and costs to calculate a net benefit. The assignment of the weights requires a management judgment regarding the relative value of avoiding curies in the case when the PC is present versus the relative value of avoiding them needlessly when there is a false alarm. Illustrative values for this judgment are shown at the bottom of columns J and K in Table 3-5. For illustration, the Core Team used the values of 10 and 1 for these column weights. These weights imply that the magnitude of the detection benefit of finding a PC and avoiding the associated curies released is roughly ten times greater than the "false-alarms costs" of needlessly avoiding the curies (and potentially rejecting a good site because of a "false positive" test result).

Finally, Column L in Table 3-5 lists the net benefits of tests. This column represents an overall figure of merit used in this report to establish priorities for early detection of potentially unsuitable site conditions. The net benefit is computed as the weighted difference between detection benefits and false-alarm costs (Cols. J minus K in Table 3-5):

Net value of test =
Weight for detection benefit x detection benefit of test −
Weight for false-alarm cost x false-alarm cost of test.

The weights for detection benefit and false-alarm cost must sum to 1.0. If they do not, then normalized weights can be computed as follows:

Weight for detection benefit = Wt1 / (Wt1 + Wt2)
Weight for false-alarm cost = Wt2 / (Wt1 + Wt2).

The relative weights Wt1 and Wt2 are assessed by management and examples are shown at the bottom of Table 3-5.

The net benefit column (Col. L) includes some negative net benefits . For example, the simplest GWTT test has detection benefits equal to 9.2×10^{-7} and false-alarm costs of 1.1×10^{-4}. The weighted difference of these is -9.2×10^{-6}. A negative net benefit can be interpreted in either of two ways. The first is the simplest: the test is apparently not worth doing, because its false-alarm costs outweigh its detection benefits.

The second interpretation is more subtle. To obtain a negative net benefit, there is an assumption that the decision maker would act in accordance with the test results, i.e., taking action to avoid the curies released, if the test detects the PC. In fact, if the net benefit is negative, such action is *not* the rational decision to make. Negative net benefits indicate that the expected consequences of taking action following the detection of a PC are worse than if no action were taken, primarily because the detection is most likely to be a false alarm. Therefore, it is better not to act given apparent detection of the PC. But if the decision rule is to take no action after a positive test result, then this is the same decision (i.e., no action) that would be taken if the test were not conducted. Thus, the test has no influence on decision making and no value in the context of avoiding curies released through early detection of potentially unsuitable site conditions. In this situation, the test is said to have zero value of information, because it does not affect the decision maker's actions. According to this interpretation, the negative entries in Col. L should be zeros. However they have been left as negative to emphasize that the test has greater potential false-alarm costs than detection benefits, based on this analysis.

The tests that fall into this category are:
- GWTT Boreholes
- GWTT, using existing core samples
- UO_2 solubility studies
- Volcanism studies related to determine the presence of a new magma body.

Of course these results change when the management value judgment on weights changes. The list above obtains when the judgment on weights is 10:1. If detection benefits and false-alarm costs are weighted equally (1:1), then *only* the following tests have positive net benefits:
- Gas flow
- Climate
- Oxidizing ground water
- Volcanism studies related to rate of events
- Reactive ground-water chemistry.

A discussion of these results is found in the following paragraphs.

Insights Regarding the Net Benefits of Testing

Figure 3-8 portrays graphically many of the issues discussed in the preceding section. The horizontal axis lists the potential concerns from Tables 3-4 and 3-5. The upper line plots the importance numbers from Col. I, Table 3-5. The units on the left axis are expected releases avoided, normalized to the EPA release limits. These units are translated to excess cancer deaths avoided on the right axis, using the ratio 1.0 on the left vertical scale to 700 excess cancer deaths over 10,000 years on the right scale.

The importance (upper) curve is essentially the same curve as in Fig. 3-2, however the italicized changes in Table 3-4 are incorporated and fewer PCs are included in the graph. Three of the PCs that have multiple levels of testing (complex geology, GWTT, and perched water) have the same importance number for all levels of testing.

The PCs in Fig. 3-8 can be grouped into three categories of roughly equal importance. As will be shown later, the benefits of testing within each group are about the same.

Group 1	Group 2	Group 3
Gas	Natural resources	Perched water
Complex geology	Climate	UO_2 solubility
	GWTT	Volcanism
	Oxidation potential (Eh)	Reactive ground water (TDS).

As was mentioned earlier, the importance curve represents the maximum possible value of testing for detecting unsuitable site conditions. "Natural Resources" has the highest importance in Groups 2 and 3, with an importance of 3×10^{-6}, which

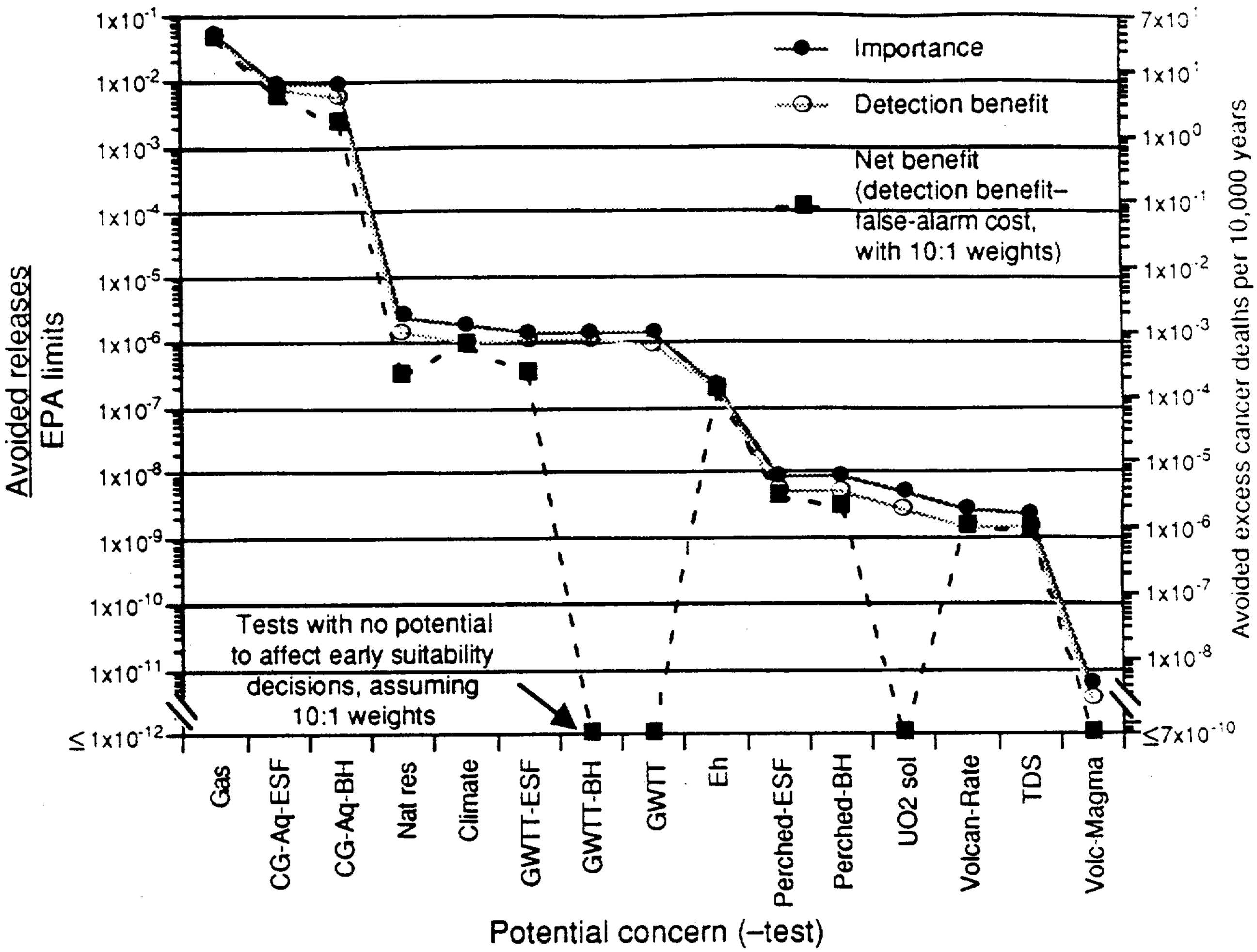

Figure 3-8. Importance and testing benefits of potential concerns. This graph shows the net benefits of early testing for the presence of the most important potential concerns (PCs), which are on the horizontal axis. Four of the PCs have different levels or types of testing evaluated. Benefits are computed assuming that, if a test detects the presence of a PC, then the potential releases and potential cancer deaths will be avoided. The upper curve indicates the importance of each PC, computed from its probability and consequences of occurrence. This curve is essentially the same as found in Fig.3-1. The middle curve indicates the detection benefit of tests, which is the importance of a PC times the probability of accurately detecting it. The proximity of the importance and detection benefit curves indicates that the importance term essentially determines the detection benefits. The lowest curve, net benefit of testing, takes into account both the probability of accurate detection and the probability of a false indication that the PC is present when, in fact, it is not present. In four cases, the "costs" of such a false alarm outweigh the detection benefits, and therefore the test is not a useful indicator of potentially unsuitable site conditions. Such comparisons of net testing benefits require a value trade-off between detection and false alarms. The results portrayed in this graph were computed using 10:1 weights for detection benefits and false-alarm costs.

corresponds to $2x10^{-3}$ excess cancer deaths avoided over 10,000 years, or one excess cancer death avoided in approximately 6 million years. This sets a *maximum* value of testing in Groups 2 and 3 that is equivalent to avoiding fewer than one onehundredth of an excess cancer death over 6 million years, using the EPA's conversion between releases and cancer deaths.

When the potential inaccuracies of testing are considered, the benefits of testing, of course, decrease. This occurs in two steps. First, the possibility of failing to detect a PC that is actually present reduces the benefits of testing, shifting them from the upper down to the middle curve in Fig. 3-8. The middle curve is a plot of the detection benefits (Col. J from Table 3-5). Second, the possibility of a false alarm when the PC is not present reduces the net benefits to the lowest curve in the figure, which is the net benefit curve (Col. L from Table 3-5).

The detection benefit curve takes into account the importance of each PC (the upper curve) and the conditional probability that a test will detect a PC that is present. This curve is close to the importance curve because most of the tests are relatively likely to detect existing PCs. In cases where there are multiple tests associated with a PC (e.g., complex geology), the detection benefit curve drops from the more accurate to the less accurate tests. The least-accurate test in the sense of detecting existing PCs is the testing program for UO_2 solubility. This has 0.5 probability of detecting the PC if it is present, and the detection benefit curve has half the benefit (2.6×10^{-9}) of the importance curve (5.1×10^{-9}). These numbers are found in Table 3-5, Cols. J and I, respectively.

Taking into account the accuracy of tests in detecting the presence of PCs does not change the conclusions reached based on the importance curve. That is, there are still three groups of test priorities, and the priorities based on the detection benefit curve are the same as priorities based on the importance curve in Fig. 3-8.

Next consider the lowest curve in Fig. 3-8, which plots the net benefit of testing defined as the benefits minus the costs. If the importance curve represents the maximum benefit of a *perfect* test, then the net benefit curve represents the maximum benefit of an *imperfect* test. In the terminology of decision analysis, the top curve is the expected value of *perfect* information, and the net benefit curve is the expected value of *imperfect* information (all measured in normalized curies released or excess cancer deaths).

In most cases, the net benefit curve is not much below the detection benefit and importance curves. This is because many tests usually have small probabilities of false alarms. In several cases, the net benefit curve drops to the bottom of the chart. These are cases where the potential false-alarms costs exceed the potential detection benefits of finding PCs that are present, and the "net benefits" are negative or zero, depending on one's interpretation of Table 3-5. (Because zero and negative numbers cannot be plotted on a logarithmic scale, the number 1×10^{-12} was was used to represent numbers less than or equal to zero.) In particular, the following tests have false-alarm costs exceeding their detection benefit, and therefore have no value according to this figure:

- GWTT borehole tests
- GWTT tests based on currently available data
- UO_2 solubility
- Volcanism test for a magma body.

The tests for which false-alarm costs exceed detection benefits depends on the management judgment regarding benefit-cost weights discussed earlier. This list of four such tests is based on a weighting ratio of 10:1. The next section shows how this list grows or contracts with different weighting ratios.

It is also useful to consider the issue of the "absolute" benefits of a test, measured in terms of avoided curies released (left-side vertical axis in Fig. 3-8) and excess cancer deaths (right-side vertical axis). If the benefits of an early test are small, then the test may not be worth its cost in time and resources. The information in Fig. 3-8 can be used to help make that type of judgment. For example, a decision maker could draw a horizontal line on Fig. 3-8 at a number of avoided curies or cancer deaths that would, in his or her judgment, clearly justify the costs of the test. If the decision maker draws the line at 10^{-5} expected releases avoided (.01 excess cancer deaths per 10,000 years), this would imply that all the tests below that line on the chart would have greater dollar costs than health and safety benefits, and they would not be worth conducting early for the purpose of detecting a potentially unsuitable site. This is a second value judgment that management needs to make, and it will be discussed further in the next section.

Step 5. Evaluate Test Priorities

The paragraphs above provide several insights about the factors that influence test priorities, such as the detection benefit, false-alarm cost, benefit-cost weights, and the minimum detection benefit required to justify the dollar costs of testing. In this section all of these factors are combined to produce a set of test priorities and associated insights regarding the effect that each factor has in determining the priorities. This section concludes with a summary of the relative advantages of each package of tests.

Test-priority groups
Figure 3-9 plots the detection benefits of tests versus their false-alarm costs. As stated in Step 4, the detection benefits are the expected releases or excess cancer deaths avoided by "finding" a PC that may be present at the site. This is plotted vertically in Fig. 3-9. (The numbers are from Col. J in Table 3-5.) The left side of the graph is scaled by avoided releases and the right side is scaled by excess cancer deaths potentially avoided in 10,000 years, for a repository with 70,000 metric tons of heavy metal. The horizontal axis plots the expected false-alarm costs from Col. K in Table 3-5, expressed as releases avoided unnecessarily.

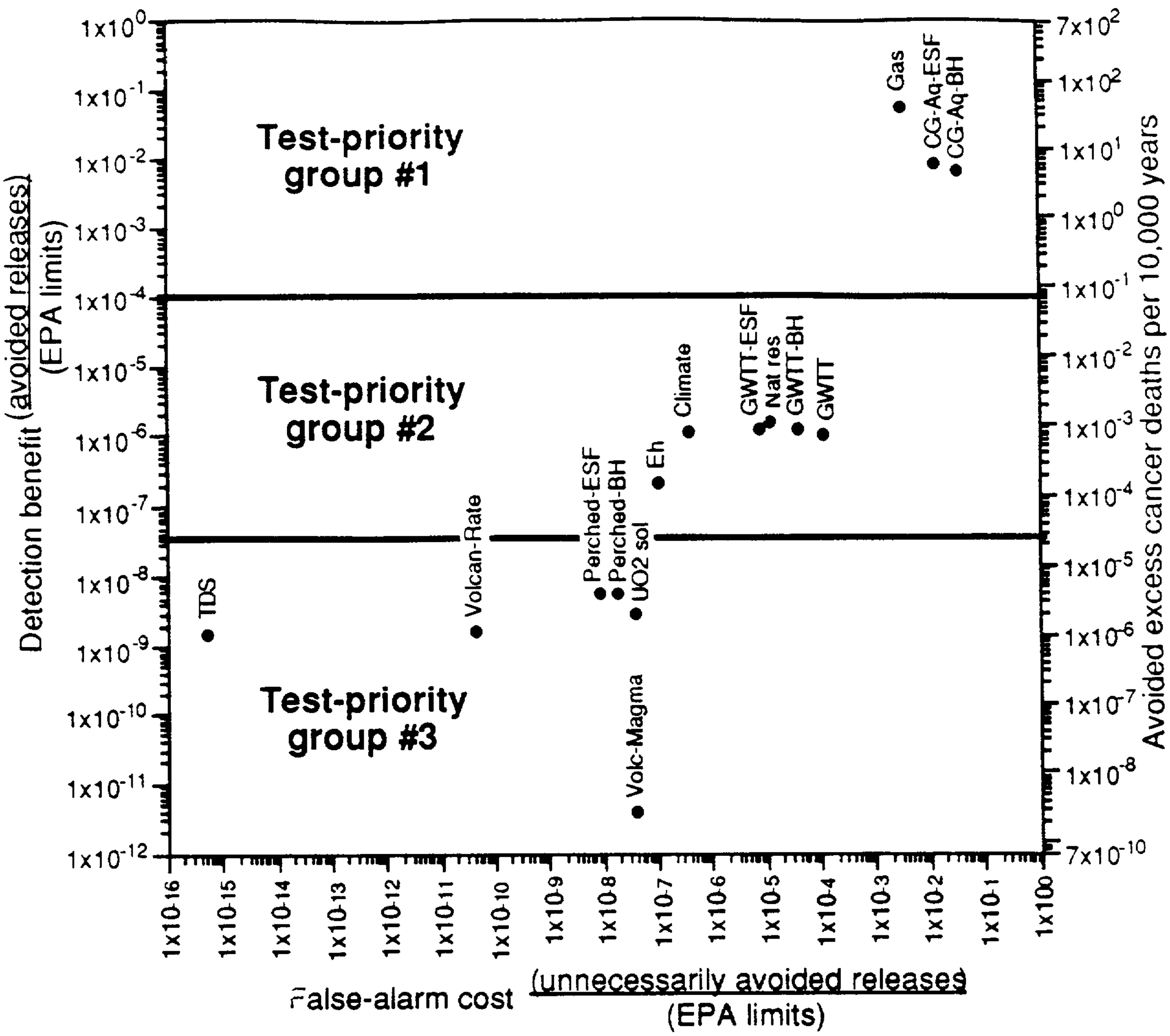

Figure 3-9. Detection benefits, false-alarm costs, and priority groups for early tests to detect potentially unsuitable site conditions. This chart plots the two primary factors in determining testing benefits for various potential concerns (PCs) and tests. These factors are detection benefits (probability of occurrence x consequences of occurrence x conditional probability of detection) and false-alarm costs (probability of non-occurrence x conditional probability of false detection x consequences). The consequences of false alarm are assumed to be proportional to the consequences of the PC if it is present at the site. There is a gap of nearly four orders of magnitude between the top three tests and the rest of the tests. Therefore, they are identified on the chart as Test-priority Group #1. Two other test-priority groups are identified. A decision maker may be able to eliminate from further consideration for early evaluation of site suitability those tests that lack sufficient detection benefits to justify the dollar costs of conducting the test. The dollar resources required for testing per excess cancer death avoided may be useful in determining which groups should be included and which should not.

Two horizontal lines divide the chart into three test-priority groups, based on detection benefits. The detection benefits of tests in any particular group are roughly the same, and they are at least a factor of 100 different from detection benefits of tests in any other group.

A decision maker can use Fig. 3-9 to decide which test-priority groups are "worth" conducting early to detect potentially unsuitable site conditions. The benefit of conducting tests decreases exponentially when one selects tests from groups that are low on the chart. This is because the PCs that those tests might detect are very unlikely or highly inconsequential. In such cases, the benefits of detecting the PC and taking subsequent action to prevent its associated releases are minimal and may not justify spending resources on conducting early tests.

One way to decide between which test groups are worth doing and which are not is to consider explicitly the dollar costs of the tests in the group. Assume a particular test costs $10 million and has detection benefit of 10^{-6} expected releases, which corresponds to 7×10^{-4} or 0.0007 avoided excess cancers over 10,000 years. Conducting that test to detect potentially unsuitable site conditions implies that spending at least $14 billion ($10 million÷.0007) is justified to avoid one excess cancer death over 10,000 years. Decision makers will ultimately judge the appropriate level of expenditures for cancer avoidance. The judgment may be explicitly related to cost per cancer avoided, or implicitly, whereby cost per cancer is not a decision criterion but is determined as a consequence of test benefits of selected tests.

Another way to determine which test-priority groups are justified is to judge first the maximum dollars to be spent to avoid one excess cancer death. Decision makers commonly make such judgments applied to other activities that affect public health and safety, such as those related to nuclear power plants, public transportation, and environmental protection. Judgments in the range $1–10 million are common in the literature for such activities (O'Riordan, et. al., 1987).

For example, if $10 million per statistical fatality were used, then the expenditures (testing costs) that would be justified in each test group are shown in Table 3-6.

Table 3-6. Illustrative justifiable testing expenditures (assuming $10 million expended to prevent a premature cancer is worthwhile)

Test-priority group	Justifiable expenditures for all tests in the group ($)
1	40,000,000.00
2	1,000.00
3	.03

Table 3-6 can be scaled to show the justified expenditures for any value per statistical fatality. If the value of preventing a cancer increases by, say, a factor of 10, the table entries are multiplied by 10. For reference, the value of preventing a single excess cancer death must increase to $10 billion (i.e., 1,000 times greater than values commonly seen in the literature) before expenditures of $1 million can be justified for conducting any or all of the tests in Test Group 2. It must be emphasized that testing and the dollar cost of testing here refers specifically to early testing to detect the presence of PCs at the site and does not apply to the overall site-characterization program. As has been pointed out several times in this report, there are many reasons for performing tests at the site other than for early evaluation of site suitability.

One can use this same approach, coupled with Fig. 3-9, to select the appropriate test group. One first determines the maximum dollars to be spent to avoid one excess cancer death. Next, multiply each number on the "avoided excess cancers" axis in Fig. 3-9 by the maximum dollar amount. Then, select the test group on the graph that contains tests with dollar costs roughly equal to the product of the maximum dollars and avoided cancers. Tests in the selected group or higher groups are "worthwhile"; tests in lower test-priority groups have detection benefits that are too low to justify testing solely for the purpose of detecting potentially unsuitable site conditions.

Trade-off Between Detection Benefits and False-alarm Costs

Using Fig. 3-9 to select which test-priority groups to conduct can help screen out tests that generate too little detection benefit to justify their cost. The next step is to examine more closely the benefits of testing *within* selected test-priority groups, based on a comparison of the detection benefits and false-alarm costs. Those tests with high false-alarm costs should be considered carefully and should not be conducted solely for the purpose of detecting potentially unsuitable site conditions.

As discussed in Step 4 above, and illustrated in Table 3-5 and Fig. 3-8, the net benefit of testing is the weighted difference between the detection benefits and the false-alarm costs of testing. The benefit-cost weights reflect a judgment of the relative significance of correctly finding a PC and avoiding a given release, compared to the significance of incorrectly finding a nonexistent PC and acting needlessly to avoid its associated releases. The weights used in the examples so far are 10:1, i.e., detection benefits are weighted ten times greater than false-alarm costs. This was indicated in Table 3-5, where these same benefit-cost weights were used to combine the detection benefits in Col. J with the false-alarm costs in Col. K.

The weighted difference between detection benefits and false-alarm costs is the difference between the vertical and horizontal coordinates of each point in Fig. 3-9. Figure 3-10 adds some diagonal lines and shading to the information from Fig. 3-9 to show where the weighted difference between detection benefits and false-alarm costs is negative (hence, where tests should be carefully scrutinized because of their potentially high false-alarm costs). Each diagonal line in Fig. 3-10 is the locus of points where the *weighted* difference between detection benefits and false-alarm costs is equal to zero. Points to the right and below the diagonal line (i.e., in its corresponding shaded area) have false-alarm costs that exceed their detection benefits, that is, the net benefit of the test is negative. The two diagonal lines and shaded areas correspond to weights of 1:1 and 10:1 and are provided for illustration only.

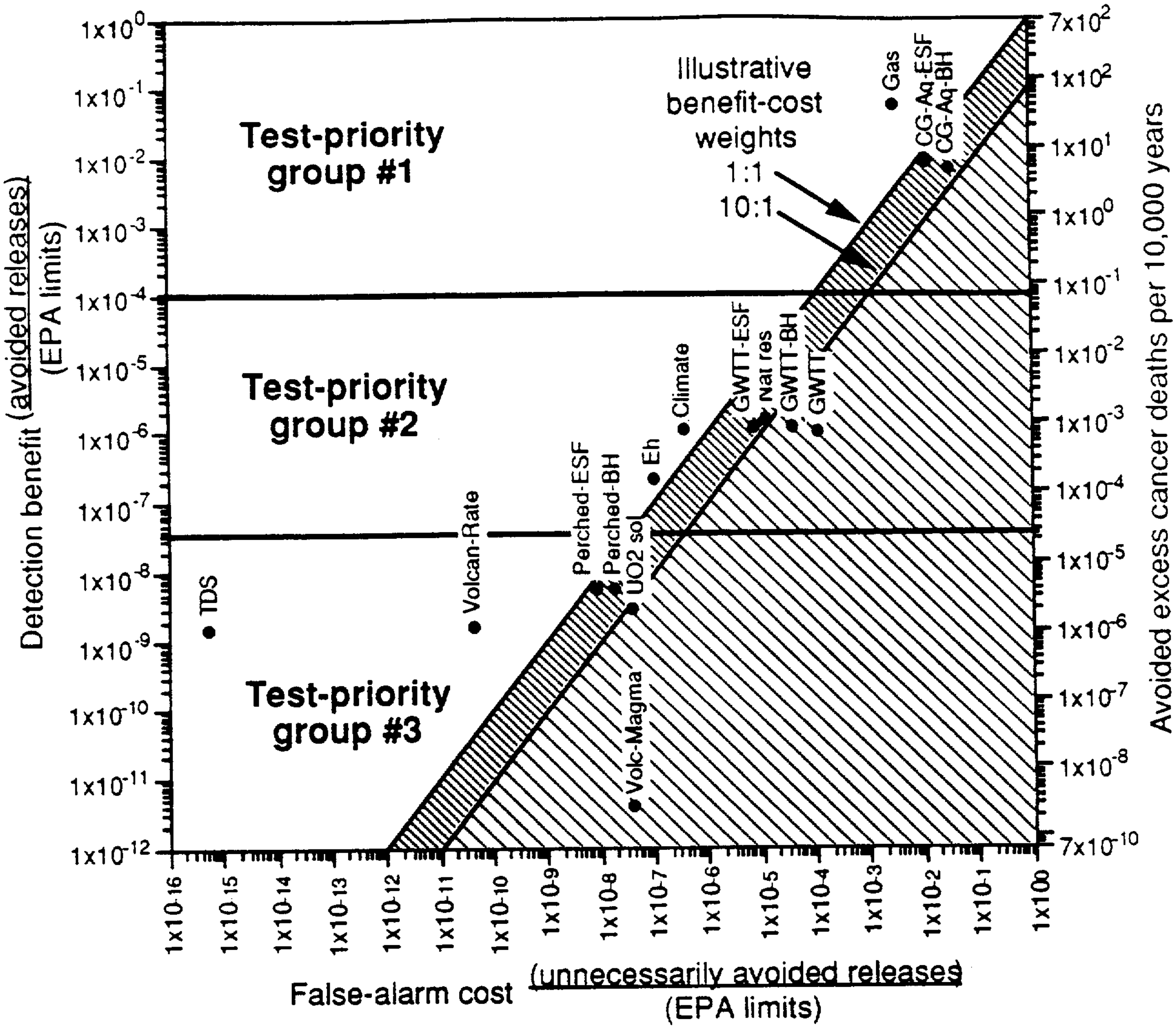

Figure 3-10. Identification of tests with excessive false-alarm costs. This figure identifies those tests in each test-priority group that have false-alarm costs that exceed their detection benefits, i.e., negative net benefits of testing. The net benefit of testing is the weighted difference between detection benefits (vertical axis) and false-alarm costs (horizontal axis). The diagonal lines represent two illustrative "benefit-cost weights," as might be determined by managers. Points in the shaded area to the right or below a given line have negative net benefits. The upper diagonal line represents a hypothetical judgment that the beneficial consequences of detecting a particular PC *equal* the undesired consequences of false detection and subsequent unnecessary action taken to avoid releases (i.e., the benefit-cost weights are 1:1). The lower diagonal, with 10:1 benefit-cost weights, assumes an accurate detection is ten times more valuable than a false detection. One uses this chart by first selecting test-priority categories where detection benefits are sufficient to justify the dollar costs of testing. Then one determines the location of the diagonal line that reflects his or her judgment on benefit-cost weights. Points above and to the left of the diagonal line represent useful tests for detecting potentially unsuitable site conditions. Points near the chosen diagonal line are sensitive to the benefit-cost weights. Tests represented by points in the shaded area to the right of the line should not be conducted solely for the purpose of detecting potentially unsuitable site conditions. Even if they are justified on other grounds, such tests are likely to result in a false alarm, and therefore a strategy should be developed for dealing with a positive result indicating the presence of a particular PC.

In each test-priority group, some of the tests fall into the shaded negative-benefit region. These tests are of questionable value as tests for detecting potentially unsuitable site conditions.

The decision maker uses Fig. 3-10 by first judging an appropriate set of benefit-cost weights, and then constructing a corresponding diagonal line. A simple way to construct the line for a given ratio of weights is to pick an appropriate "anchor" point on the horizontal (bottom) scale, and then draw a line parallel to the two shown in Fig. 3-10. For example, if the weights are 100:1, the anchor on the vertical axis is at 10^{-10}, which is a factor of 100 times the anchor (10^{-12}) for the 1:1 line. If weights of 1:100 were chosen, the anchor would be at 10^{-14}, or .01 times the anchor for 1:1 weights.

As mentioned, tests represented by points to the right and below the constructed line cannot be justified solely for the purpose of detecting potentially unsuitable site conditions. Even if they are undertaken for "other reasons," one must bear in mind that they are very likely to give a false alarm, and one should develop a strategy for dealing with positive test results that may falsely indicate the presence of a PC.

Three other insights can be derived from Fig. 3-10. First, the points appear to fall in a region parallel to the two diagonal lines from the lower left to upper right corners of the graph. This is a consequence of the assumption that the detection benefits and false-alarm costs are both proportional to the consequences of the PCs, measured in terms of curies released. High-consequence PCs appear at the upper right corner of the graph.

Second, the benefit-cost weights might be different for different tests. For example, the relative consequences of accurate detections and false alarms might be greater for ground-water travel time than for, say, perched water. In such cases a single diagonal line and shaded area cannot be used for both types of tests. However the chart is still useful; one simply constructs a family of parallel diagonal lines, identifies the appropriate line for each test, and then excludes or includes the test depending on its position relative to its associated diagonal line.

Third, tests that fall on the left side of the decision maker's diagonal line have positive net benefit of testing. Those tests that fall near the decision maker's diagonal line (or just to the right of the line) are tests whose overall evaluation depends critically on the value judgment regarding weights. In Fig. 3-10, the following tests fall in between the two illustrative diagonal lines, and are therefore sensitive to the judgment regarding weights:

- Complex geology
- Natural resources
- Climate
- GWTT
- Oxidation potential (Eh)
- Perched water
- UO_2 solubility.

Those tests that are not located near the diagonal lines are less sensitive to the value judgment on weights and include the tests for the following PCs:

- Gas flow radionuclide
- Reactive ground-water chemistry (TDS)
- Volcanism tests related to rate of volcanic events
- Volcanism tests related to new magma body.

Once the appropriate test-priority groups are selected and the location of the diagonal line representing the benefit-cost weights is determined, then the chart is easy to interpret: priority is to be given to the tests that plot above and to the left of the lines within each selected test-priority group.. Furthermore, one can determine from the chart how "close" various tests are to the dividing lines between "test" and "don't test" regions.

Sensitivity Analysis for Ground-water Travel Time

All of the analyses discussed to this point have focused on post-emplacement releases of curies over 10,000 years. There are other criteria for judging the waste-isolation capabilities of Yucca Mountain. For example, the analysis of PC #6, Ground-water Travel Time (GWTT), was tied to expected curies released if the GWTT is less than 1,000 years. The actual GWTT performance objective set by the Nuclear Regulatory Commission (NRC) is that GWTT should be greater than 1,000 years and does not refer to potential releases. If the NRC performance objective is considered equivalent to meeting the EPA limits on total-system performance, then the evaluation of the three levels of GWTT tests in Fig. 3-10 changes somewhat.

Figure 3-11 shows how the detection benefits of GWTT tests increase if failing to meet the GWTT performance objective is equated with failing to meet the EPA limits. In previous calculations, the consequences of failure to meet the GWTT objective were about a factor of 600 below the consequences of failure to meet the EPA release limits. This can be seen in Table 3-2, where the product (C–1)xD for the GWTT PC is 1.7×10^{-3}, or approximately, 1/600. Multiplying the expected consequences (Col. E from Table 3-4) by 600 approximately equates the value of the GWTT objective and the EPA containment requirement.

The effect of treating the GWTT performance objective as having equivalent consequences to the EPA release limits is to raise the GWTT tests above the illustrative boundary between Test-priority Groups #1 and #2. However, both detection benefits and false-alarm costs increase, so the values one chooses for benefit-cost weights remain critical to the determination of priorities. Only if the benefit-cost weights are 10:1 do any of the GWTT tests have benefits exceeding false-alarm costs, because the probability of a false alarm is so high. The GWTT-ESF tests are indicated to have positive benefit at 10:1 benefit-cost weights; however, these are not tests that are likely to be performed early during site characterization.

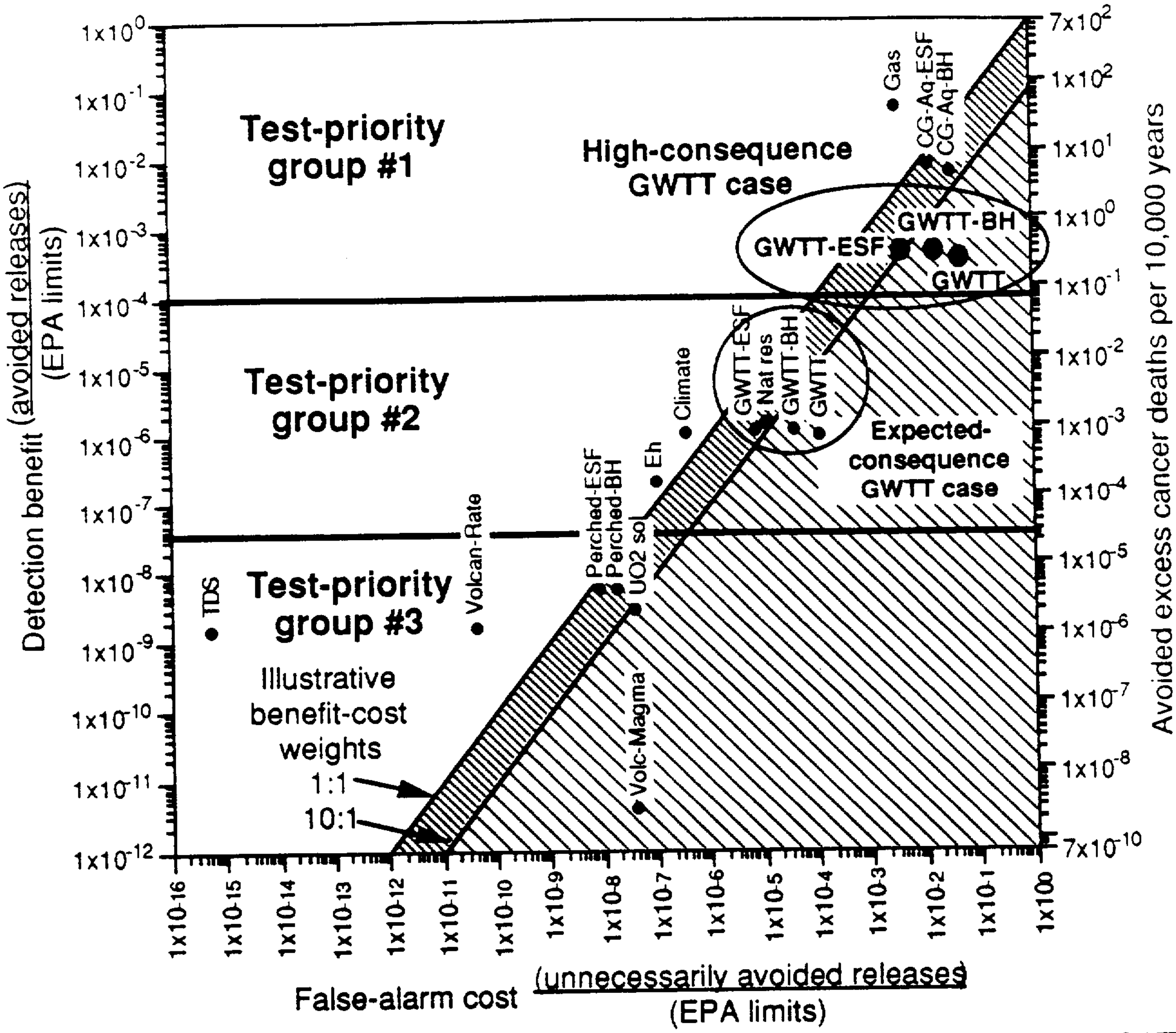

Figure 3-11. Sensitivity of ground-water travel time (GWTT) tests to interpretation of the 1,000-year GWTT performance objective. If one considers failure to meet the 1,000-year GWTT objective as having consequences equal to violating the EPA release limits, then the importance and testing benefits associated with GWTT increase by a factor of 600. The bold dots represent GWTT tests in this high-consequence sensitivity case. The smaller dots represent the expected-consequence case, based on the curies released. Increasing the consequences of violating the 1,000-year GWTT objective increases the detection benefit, and, depending where the dividing line is drawn, could put GWTT in Test-priority Group #1. At the same time, however, false-alarm costs increase. Thus, the GWTT tests are still of questionable value for detecting unsuitable conditions due to the potential for false alarms, and their net benefit depends critically on the benefit-cost weights.

Summary of Analysis Results for each Potential Concern

The following paragraphs summarize the results for each potential concern and each testing package. The conclusions drawn are based on the benefits of conducting tests that may detect the presence of potential concerns. The discussion is organized using the groups of tests identified earlier.

Test-priority group 1. High-priority tests

- *Gas flow.* This potential concern has the highest overall benefit of testing and is relatively insensitive to management judgments regarding benefit-cost weights.

- *Complex Geology.* Both the borehole and ESF levels of testing for complex geology have high benefits, but both are sensitive to the judgment on benefit-cost weights. For example, both testing levels are worthwhile only if the ratio of weights is about 10:1 or greater.

Test-priority group 2. Middle-priority tests (low detection benefits)

- *Natural resources.* This PC is in Group #2 because it has low probability of occurrence (.002) and low consequences if it is present (.001). The benefits of testing barely exceed false-alarm costs even if the benefit-cost weights are 10:1.

- *Climate effects.* This PC has relatively high probability of occurrence (.25) but very low consequences if it occurs ($7x10^{-6}$). The value of the test to detect this concern is relatively insensitive to the benefit-cost weighting factor.

- *GWTT.* This PC has low probability of being present (.0015) and low consequences ($9x10^{-4}$) if present, with respect to possible radionuclide releases. Three levels of testing were considered, but only testing in the ESF has net positive benefit even for 10:1 benefit-cost weights. This is because the probability of a false alarm is high. However, placing reliance on ESF testing may not provide *early* detection of potentially unsuitable site conditions.

  If failing the 1,000 year GWTT performance objective is judged to have waste-isolation consequences equal to violating the EPA release limits, then both the detection benefit and false-alarm cost increase by a factor of 600. Again, the ESF testing has positive net benefit only for 10:1 benefit-cost weights.

- *Oxidizing ground water (Eh).* The PC has high probability of occurrence (.4) but very low consequences ($6x10^{-7}$). The test for oxidizing ground water is relatively insensitive to the judgment on weights.

Test-priority group 3. Low-priority tests (very low detection benefits)

- *Perched water*. The PC has moderate probability of being present (.03) but very low consequences ($3x10^{-7}$). Its importance is $8x10^{-9}$, which, like the other tests in this group, is likely to place this PC below the minimum detection benefit required to justify the dollar costs of tests. Two levels of testing were considered, but both have about the same net benefit of testing. Benefits outweigh false-alarm costs as long as the benefit-cost weight ratio is greater than 1:1.

- *UO_2 solubility*. The PC has low probability of being present (.005) and very low consequences ($1x10^{-6}$). Its importance is $5x10^{-9}$. The ratio of benefit-cost weights must be greater than 10:1 before this should be considered.

- *Volcanism*. Two types of investigations of PCs were considered: those to determine rates of volcanic events and those to detect the presence of a new magma body. Low probability ($2x10^{-5}$) was assigned to the rate of volcanic activity exceeding the assessment threshold, and, even if the rate is above the threshold, it still has low consequences ($1x10^{-4}$). The detection benefit of the PC is $2x10^{-9}$, which is very low compared to other PCs. However, testing to determine the rate of volcanic activity is insensitive to the benefit-cost weighting factors.

  The concern about a new magma body has low probability of being present ($4x10^{-5}$) and very low consequences ($1x10^{-7}$). The importance is only $6x10^{-12}$, and the false-alarm costs far outweigh the benefits of the test to determine if a new magma body exist.

- *Reactive ground water (TDS)*. This PC has low probability of occurrence ($4x10^{-4}$) and very low consequences ($6x10^{-6}$). Its importance is $2x10^{-9}$, which like the others in this category, is very low. The test for reactive ground water is highly insensitive to the weighting factors used; benefits outweigh false-alarm costs.

4. Conclusions and Recommendations

Before summarizing conclusions and presenting recommendations, it is useful to review briefly the project objective and test prioritization criteria. This chapter begins with that review. Then analytic results are summarized, along with several insights gleaned from the assessment and analysis. The chapter ends with recommendations for test priorities and for further use of the analytic method and assessed information.

Review of Project Objective and Prioritization Criteria

Project Objective

The objective of this study is to identify tests that could be conducted early during site characterization to detect potentially unsuitable site conditions, if they exist at the site. The study analyzed 32 potential concerns (PCs) in detail, which were derived primarily from the potentially adverse and disqualifying conditions listed in 10 CFR Part 960 and the potentially adverse conditions in 10 CFR Part 60.122(c). This analysis is intended to answer two questions: (1) "Which of these concerns have the greatest potential for rendering the site unsuitable with respect to possible postclosure radionuclide releases to the accessible environment?" and (2) "Which tests are most likely to provide accurate detection of these concerns if they are present at the site?"

Criteria for Evaluating Tests

Testing benefits in this analysis are expressed in terms of the consequences for the waste-isolation capabilities of the site. Tests are judged beneficial when they can detect potential concerns correctly and thereby allow decision makers to avoid the detrimental effects that may be caused by those concerns. Tests are presumed to be not beneficial when they have an appreciable probability of leading to false alarms, that is, to indications that PCs are present when, in fact, they are not. Postclosure radionuclide release to the accessible environment over the next 10,000 years was used as the measure of all potentially detrimental effects on waste isolation and, therefore, on public health and safety. All releases were normalized using the ratio: cumulative curies released to the accessible environment over 10,000 years divided by the EPA limits on releases.

This analysis found that very few of the potential concerns are of sufficient importance to merit early investigation for evaluating site suitability. This conclusion results either because the PC has a very small probability of occurring at the site or because its potential impact on waste isolation is deemed to be small.

This conclusion does not mean that these concerns should not be investigated for other reasons, such as:

- Building scientific consensus about the evaluation of site suitability
- Gathering information for repository design and construction
- Providing ancillary information required for a license application
- Providing baseline data for long-duration performance-confirmation tests during and following repository construction.

A second conclusion of this analysis is that there is a high potential for false alarms from any test, regardless of the reason for testing. Decision makers need to be cognizant of the possibly significant consequences of false alarms as they plan the testing program.

Analysis Results and Insights

Importance of Potential Concerns

The analysis in Steps 1 and 2 identified and evaluated the importance of 32 PCs. The importance of each PC can be interpreted as the expected value of *perfect* information about its presence or absence at the site (i.e., 100-percent test accuracy). This sets an upper bound on the value of any practical testing activities that are aimed at detecting the presence of PCs.

The importance of the 32 PCs spans a range of 14 orders of magnitude. Thus, on the basis of the expected effects on waste isolation, some of the PCs are much more important than others. Using the EPA conversion factor between releases and excess cancer deaths, the releases for PCs range from a high of 39 to a low of 1.3×10^{-12} expected excess cancer deaths over 10,000 years (i.e., about one every 250 to 7×10^{15} years). The EPA limit on site performance for this repository allows 700 excess cancer deaths per 10,000 years (one every 14 years).

The set of PCs considered in this study were divided into three Importance Groups:

1. High relative importance
   Three PCs that relate to releases of gas-phase radionuclides (specifically, carbon-14) and to complex site geology that affects gaseous and aqueous releases and could significantly complicate modeling site performance
2. Medium relative importance
   Eleven PCs that relate to human intrusion, ground-water travel time, geochemical conditions in the host rock, perched water, and igneous activity at the site
3. Low relative importance
   All 18 remaining PCs shown in Fig. 3-1

Because the PCs in Group 3 are unlikely to affect waste isolation, they were judged not to require further evaluation in this study. The 14 PCs included in Groups 1 and 2 were carried forward to the test-accuracy part of the analysis.

Test Accuracy
Because there was some duplication of concerns among the PCs evaluated for testing , the original list of 14 PCs was narrowed to a list of 10 PCs. A total of 15 packages of tests for these 10 categories of PCs were evaluated with regard to their accuracy for detecting the presence of a PC.

The assessed accuracy of these test packages also varied, but not as widely as the importance. The conditional probabilities of true positives (finding the PCs given that they are present) ranged from 50 to 98 percent. The conditional probabilities of false alarms ("finding" the PC when, in fact, it is not present) ranged from nearly zero to 29 percent. Several test packages are quite inaccurate, thus reducing their net benefits. *In general, tests that have a relatively high probability of false alarm are undesirable when they are conducted to search for a PC that is unlikely to be present.*

Tests should not be judged solely on the basis of their conditional probabilities of true and false positive results. There are at least three reasons for this. First, these two probabilities can be manipulated by changing the definition of a "positive" test outcome. In particular, the conditional probability of a true positive can be made to approach 100 percent, which means that the test will *always* detect a PC. At the same time, however, the probability of a *false* alarm will be very high. Second, these two probabilities do not account for the importance (the likelihood and consequences) of the PC that the tests are intended to detect. Third, these two probabilities alone fail to take account of the relative value of a true-positive result compared to the relative (negative) value of a false-positive result. This analysis of testing priorities considers these additional factors.

Test Priorities
The net benefits of testing are defined in this report as the weighted difference between the detection benefits and false-alarm costs. Detection benefits, the expected radionuclide release that could be avoided if the PC were both present and detected by the test, ranged from 4×10^{-12} to .05 times the EPA release limits. Expressed as excess cancer deaths avoided, this range is 2.5×10^{-9} to 36 deaths over 10,000 years (or one excess cancer death avoided roughly every 280 to 4×10^{12} years). False-alarm costs, expressed as radionuclide releases *unnecessarily* avoided because of an erroneous belief that the PC is present, ranged from 5×10^{-16} to .03, or 3×10^{-13} to 23 cancer deaths that would not have occurred in actuality (but, nevertheless, for which funds would have been expended for unnecessary mitigation measures or the site would have been abandoned unnecessarily to "prevent" them). A value judgment regarding relative weights between the benefits and false-alarms costs is required to produce an overall net benefit of testing.

On the basis of the assessed detection benefits and false-alarm costs shown in Fig. 3-10, the PCs and their associated tests were grouped into three test-priority groups. The tests in each group have similar detection benefits; those in different groups have detection benefits differing by at least two orders of magnitude in radionuclide releases.

1. High-priority tests
   Tests for gas-flow radionuclide transport above the repository and, depending on the value judgment on benefit-cost weights, the tests that address complex geology related to aqueous-phase radionuclide releases

2. Middle-priority tests
   Tests for climate changes, oxidation potential of water in the host-rock, and, depending on the benefit-cost weights, the ground-water travel time, natural-resources, and direct-intrusion tests.

3. Low-priority tests
   Tests for reactive ground wat[illegible] -he host rock, rate of volcanism and, depending on the benefit-cost [illegible] .its, perched water and UO_2 solubility.

A sensitivity analysis of the value of testing ground-water travel time was conducted. Expected ground-water travel time greater than 1,000 years is specified as a performance objective by the Nuclear Regulatory Commission. If the consequences of not meeting this objective are set equal to the consequences of exceeding the EPA release limits (i.e., cumulative releases equal to 1.0 times the EPA limits), then the detection benefits of the ground-water travel time tests increase by a factor of 600. In this case, the equivalent benefits are on the order of .001 times the EPA release limit or 0.5 cancer deaths avoided over 10,000 years (one excess death every 20,000 years). However, the false-alarm costs increase proportionally, which makes the net benefits of early tests to determine if the GWTT is less than 1,000 years negative, unless relative benefit-cost weights of 100:1 are assigned.

Value Judgment on Minimum Detection Benefits
A critical judgment is, "What minimum level of detection benefit is required to justify the costs of early testing for unsuitable site conditions?" Below this level, testing may have value in reducing potential releases, but that value is exceeded by the dollar cost of the tests.

Based on this concept of a minimum detection benefit, three distinct test-priority groups were identified:

- Test-priority Group #1. Tests in this group have detection benefits expected to avoid at least one excess cancer death every 280 to 2,500 years.

- Test-priority Group #2. Tests in this group have detection benefits expected to avoid at least one excess cancer death every 10 million to 70 million years.

- Test-priority Group #3. Tests in this group have detection benefits expected to avoid at least one excess cancer death every 3 billion to 4 trillion years. Such levels imply value judgments that are clearly below the levels of expenditures to protect human life that are characteristic of other social decisions. (cf. O'Riordan, et. al., 1987.)

It must be emphasized that testing here refers specifically to early testing to detect the presence of PCs at the site and does not apply to the overall site-characterization program. As has been pointed out several times in this report, there are many reasons for performing tests at the site other than for early evaluation of site suitability.

Summary of Conclusions

The major conclusions of this study are listed below:

1. *Importance of Potential Concerns*
From the perspective of potential effects on waste isolation, some of the PCs are much more important than others. Three PCs have greater potential contribution to radionuclide releases than all others by at least a factor of 200; these include "Gas flow radionuclide," "Complex geology-gaseous," and "Complex geology-aqueous." Among these three PCs, the highest expected contribution to curies released over 10,000 years is .06 times the EPA limits.

2. *Test Accuracy*
Test accuracy was assessed for 15 test packages associated with the ten most important PCs. Test accuracy ranged from 50 to 98 percent probability of detecting the PC if it is present. False-alarm probabilities ranged from nearly zero to 29 percent. Because the probabilities of true and false positive are coupled for a particular test (and one or the other can be made arbitrarily high), test accuracy alone is not a good measure for prioritizing tests. The probability that the PC is present and the consequences if it is present also need to be taken into account.

3. *Detection Benefits*
Detection benefits measure the expected contribution of a test for detecting a PC if it is present and thereby allows action to be taken to prevent the possible consequences of the PC. Detection benefits for the 15 evaluated tests ranged from .05 to $4x10^{-12}$ times the EPA release limits. Expressed as avoided cancer deaths, this is roughly one excess death avoided every 250 to $3x10^{12}$ years, respectively.

4. *False-alarm Costs*
False-alarm costs also varied substantially: from .01 to $5x10^{-16}$ expected releases, or 8 to $3x10^{-13}$ cancer deaths, for which time and resources would have been expended unnecessarily, either for mitigation measures or for abandoning the site. For this reason some PCs may have false-alarm costs associated with early testing that may exceed their detection benefits.

5. Testing priorities

The tests of highest priority are those for gas flow (carbon-14 release) above the repository and, possibly, tests that address complex geology related to aqueous-phase radionuclide releases, depending on the value judgment on benefit-cost weights. These tests in Test-priority Group #1 have the potential to avoid one excess cancer death roughly every 280 to 2,500 years. The tests in Test-priority Group #2 contribute to avoiding only one excess cancer death every 10 million to 70 million years.

Tests for complex geology are worthwhile if the benefit-cost weighting is at least 10:1 (i.e., the curies avoided through early detection are worth at least 10 times more than the costs associated, with action taken to unnecessarily avoid releases due to false alarm).

6. Ground-water Travel Time Sensitivity

The consequences of the PC regarding GWTT were related to cumulative releases for this study; whereas the criterion that GWTT be less than 1,000 years is a separate performance objective. If the consequences of violating the GWTT performance objective are set equal to the consequences of violating the EPA radionuclide release limits, then the detection benefits for GWTT tests increase by a factor of 600. This translates into a detection benefit of avoiding one excess cancer death over 20,000 years. However, the false-alarm costs increase proportionally, which makes the net benefits negative for tests to determine whether the GWTT is less than 1,000 years, unless relative benefit-cost weights greater than 100:1 are assigned.

7. Value Judgment on Minimum Detection Benefits

A judgment regarding the minimum level of detection benefit required to justify the dollar costs of testing is required before one can judge whether to conduct any tests in a particular test-priority group. Because there are only three distinct groups of tests, one needs only to choose among three minimum detection benefit levels. The minimum detection benefit in Test-priority Groups #1, #2, and #3 can be expressed as avoiding at least one excess cancer death every: 2,500 years, 70 million years, or 4 trillion years, respectively.

8. Benefit-cost Weights

A second value judgment is needed to determine whether tests with high false-alarm costs can be justified. If the weights for detection benefits are judged to be equal to those for false-alarm costs, then the costs of many tests outweigh their benefits. This is because the tests are investigating unlikely and/or inconsequential PCs.

Recommendations

Based on the results of this analysis, the authors developed a set of recommendations in several topical areas, which include:

- Assessment of management value judgments
- Priorities for early tests
- Analyses of related issues
- Completion of the Phase II analysis
- Potential use of results in site-suitability determinations
- Further application of the approach to revise and update test priorities during site characterization.

1. Assess Management Value Judgments

Assessment of two types of value judgments by management personnel is required in order to set initial priorities on early testing to detect potentially unsuitable site conditions. The two types of value judgments are:

- Minimum detection benefits required to justify the dollar costs of testing to detect potentially unsuitable site conditions
- Relative benefit-cost weights of correct and false detection of PCs.

Once these judgments are made, Fig. 3-10 gives a clear indication of priorities for tests to detect PCs.

Such judgments can be made directly, as discussed in this report. Or, some additional modeling and assessment can be carried out in order to incorporate additional information and produce a more defensible set of value judgments. For example, some of the factors that could be considered in an assessment of the minimum detection benefit could include:

- The value of avoiding excess cancer deaths
- The dollar cost of mitigation measures
- The dollar costs of conducting the proposed tests.

A value model to assist in the assessment of the benefit-cost weights could include some of these same assessments, in particular those related to the benefit of detecting PCs and potentially avoiding excess cancer deaths. The assessment for benefit-cost weights could also include the following factors related to the costs of false alarms:

- The implications for cost and radionuclide release if the Yucca Mountain site were abandoned unnecessarily and an alternative site or alternative "back end" of the nuclear fuel cycle were developed
- The implications for dollar cost and radionuclide release of leaving spent fuel at reactors while another site or option is sought
- The likelihood that the site will be abandoned for other reasons.

Such assessments and value models would help clarify the factors influencing the minimum detection benefit and benefit-cost weights and the implications of these difficult value judgments. However, regardless of whether such modeling is undertaken or the judgments are made directly, these two judgments are essential to determining which tests are worth conducting.

2. Set Priorities for Early Tests

Once the two value judgments are made, attention can be focused on deciding which specific tests should be conducted early during site characterization. Because the potential for releases is highest for gas-phase carbon-14 (C-14), this concern received highest priority in the evaluation. Although the assessment team identified tests that could be applied to gas-flow time above the repository and potential chemical retardation of C-14 transport, there is no testing program that is directed specifically to C-14 release and transport from the repository. Consequently, it is recommended that a *strategy* be developed for addressing potential C-14 releases. Although not addressed in this study, various other options may be available and should be considered when developing a strategy for C-14 releases. Some of these options include:

- Conducting site tests to evaluate the potential for rapid transport of C-14 to the accessible environment if it escapes from the waste containers
- Testing the waste form and cladding to determine the amount of C-14 expected in the rapid-release fraction
- Venting the waste before emplacement
- Reviewing regulatory requirements regarding C-14 releases and consideration of rule changes.

The tests for air flow (air permeability) and C-14 retardation were planned to support other objectives, primarily the characterization of hydrologic features of the unsaturated zone above the repository. Such tests could be given high priority if site testing is a part of the "C-14 strategy." However, it is recommended that a testing strategy explicitly focused on C-14 transport factors be developed before assigning high priority to the currently-identified applicable tests. (Additional information can be found in Appendix D.)

Tests for complex geology could be assigned high priority, depending on a clarification of the relation between complex geology and modeling accuracy and on the management value judgment about benefit-cost weights (Recommendation 1 above). Specific tests for early evaluation of complex geology include:

- Vertical borehole investigation of the Ghost Dance and Solitario Canyon faults
- Potentiometric-level evaluation to investigate the steep hydraulic-gradient zone north of the site
- One to three boreholes from the systematic drilling program that are independent of specific features in order to provide areal control.

These tests were evaluated together, as a package. No prioritization was evaluated or implied for specific tests in the package. (Additional information can be found in Appendix D.)

This analysis does not support priority testing for other PCs for the purpose of early detection of potentially unsuitable site conditions. While there may be other reasons for giving high priority to such tests, one needs to be aware of the relatively high likelihood of these tests to yield *false* indications of unsuitable site conditions.

3. *Analyze Preclosure and other Site-Suitability Issues*
The authors recommend expanding the scope of the analysis to address preclosure or other site-suitability issues not addressed by this analysis. There may be good reasons among those issues to justify early testing of the site (e.g., seismic concerns related to preclosure operations).

4. *Complete Phase II Assessment and Analysis*
Several possible extensions of the assessment and analysis should be considered in Phase II. The number and diversity of workshop participants could be expanded to include a broader range of experts on individual PCs, possibly including experts external to the current program. The criteria for evaluating the importance of PCs could be expanded from the current postclosure total-system-performance criteria to include: preclosure health and safety, ease and cost of construction, environmental, socioeconomic, and transportation impacts, and postclosure subsystem performance. Similarly, the criteria for prioritizing tests could be expanded to include the dollar cost of tests and measures related to the "other reasons" for testing listed earlier.

In addition, expansion of the assessment and modeling should be considered. A dominant factor influencing the conclusions of this analysis is the expected consequences for waste isolation if potential concerns are present. These consequences are based on expert judgment and were difficult to assess. A simplified total-system-performance model for calculating those consequences would enhance the credibility, clarity, defensibility, and future utility of this analytic approach. Further, the inputs to a total-system-performance model would be assessed at a lower level of detail, compared to Phase I assessments. This would make the assessments easier for the experts, especially in cases where several interrelated factors were considered simultaneously in the Phase I assessments. The authors recommend that managers consider and decide which of these expansions of the analysis will most enhance its usefulness as a management tool.

5. *Use of Results in Site-Suitability Determinations*
The analytic method, assessments, and numerical results of this analysis can provide useful information and methods to the process being considered for early evaluation of site suitability. For example, the importance and testing assessments yield insight on what might be learned in the initial phases of testing. This information bears directly on the "lower-level" and "higher-level" findings required by 10 CFR Part 960. However, to address suitability issues comprehensively, the scope of the analysis would need to be broadened, as discussed above, to incorporate preclosure and other related issues. The method developed here for prioritizing testing embodies a simplified analysis of three alternatives: continue testing, apply for a license, or abandon the site. This structure is compatible with and directly applicable to factors in site-suitability decisions that consider the net benefits of these alternatives.

6. Apply Method to Reprioritize Tests During Site Characterization
The analytic method used in Phase I can be extended and reapplied at any point during site characterization. The Phase II model would enhance such applications, but the procedure would be similar with or without the Phase II model. Phase I established a foundation for future assessment and analysis, and only changes to that foundation will be required in future applications.

Similarly, the assessments obtained in this analysis can be used to update assessments of the probability that various PCs are present, given the results of early testing. This is especially important if tests are conducted that have high probabilities of false alarms, as do many of the tests analyzed in this report.

In fact, an important issue for decision makers to face is, "How should one treat results from tests that are conducted for reasons unrelated to early determination of potentially unsuitable site conditions but whose results show that potential concerns may be present." According to the assessments in this study, these test results may well be false alarms. In summary, the method developed here provides a systematic and defensible approach that could be used for updating assessments with new information, drawing inferences, and making suitability or testing decisions based on test results.

References

Apostolakis, G., 1990, The Concept of Probability in Safety Assessments of Technological Systems, Science, Vol. 250, p 1359-1364.

Bayes, Thomas, 1763, Essay Toward Solving a Problem in the Doctrine of Chances, Biometrika, Vol. 45, original in 1763, reprinted in 1958, pp 293-315.

Bonano, E.J., S.C. Hora, R.L. Keeney, D. von Winterfeldt, 1990, Elicitation and Use of Expert Judgment in Performance Assessment for High-Level Radioactive Waste Repositories, NUREG/CR-5411 SAND89-1821, Sandia National Laboratories, Albuquerque, NM.

Brown, R. V., 1987, Decision Analytic Tools in Government, in Karen B. Levitan (ed.), Government Infostructures, Greenwood Press, Westport, Ct.

de Finetti, B., La Prevision, 1937, Les Logiques, es Sources Subjectives, Annales de L'Institute Henri Poincaire, Vol. 7, pp 1-68.

DOE, U.S., 1988, Report to Congress on Reassessment of the Civilian Radioactive Waste Management Program, U.S. Department of Energy, Office of Civilian Radioactive Waste Management, June 1986.

DOE, U.S., 1988 (Dec.), Site Characterization Plan, U.S. Department of Energy, Office of Civilian Radioactive Waste Management, DOE/RW-0199.

Holloway, C.A., 1979, Decision Making Under Uncertainty: Models and Choices, Prentice-Hall, Inc., New Jersey, 522p.

Howard, R. and J. Matheson, eds., 1984, Readings on the Principles and Applications of Decision Analysis, 2 volumes, Strategic Decisions Group, Menlo Park, Ca.

Judd, B., and S. Weissenberger, 1982, A Systematic Approach to Nuclear Safeguards Decision Making, Management Science, Vol. 28, No. 3, March 1982.

Keeney, R., 1980, Siting Energy Facilities, Academic Press, New York.

McNamee, P. and J. Celona, 1987, Decision Analysis for the Professional with Supertree, The Scientific Press, Redwood City, CA.

Merkhofer, M. L., 1987, Quantifying Judgmental Uncertainty: Methodology, Experiences, and Insights, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-17, pp 741-752.

O'Riordan, T., Kemp, R., and Purdue, H. M., 1987, On Weighing Gains and Investments at the Margin of Risk Regulation, Risk Analysis, 7, pp. 361-370.

Raiffa, H. and R. Schlaifer, 1962, Applied Statistical Decision Theory, Harvard Unversity Graduate School of Business Administration, Boston, Mass.

Savage, Leonard J., 1954, The Foundations of Statistics, John Wiley & Sons, New York, (reprinted by Dover Press in 1972).

Spetzler, C. and C. A Staël von Holstein, 1975, Probability Encoding in Decision Analysis, Management Science, Vol. 22, No. 3, November 1975.

Tversky, A., and D. Kahneman, 1974, Judgment Under Uncertainty: Heuristics and Biases, Science, Vol. 185.

10 CFR Part 60, (Code of Federal Regulations), 1987, Title 10, Energy, Part 60, Disposal of High-Level Radioactive Wastes in Geologic Repositories, U.S. Government Printing Office, Washington, D.C., pp. 627-658.

10 CFR Part 960, (Code of Federal Regulations), 1987, Title 10, Energy, Part 960, General Guidlines for the Recommendation of Sites for Nuclear Waste Repositories, U.S. Government Printing Office, Washington, D.C., pp. 518-551.

40 CFR Part 191, (Code of Federal Regulations), 1986, Title 40, Protection of the Environment, Part 191, Environmental Radiation Protection Standards for Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes, U.S. Government Printing Office, Washington, D.C., pp. 7-16.