Draft Feb 2000; @pal99; 6 figs

MANAGING RESEARCH TO EVALUATE IF A RISK IS ACCEPTABLE: APPLIED TO NUCLEAR WASTE SITING

Rex V. Brown

Research Professor of Public Policy
George Mason University,
Fairfax, Va. 22030-4444
703-620-2443 (fax -2839)
rbrown@gmu.edu

ABSTRACT

Evaluating whether a potentially hazardous facility is safe enough can cost much unnecessary time, resources and even public embarrassment, unless there is firm and defensible guidance on how to manage the supporting research. This requires a test of acceptable risk and some way of assessing whether the test has been passed—-or will be passed after research.

A topical example is how to control the massive research effort being undertaken to determine whether a proposed nuclear waste site should be approved. According to regulations, radioactive release from an acceptable site should be below some limit with 90% probability. The difficulty with this test is that, even if a site passes it *now* based on current evidence, there is no assurance that the site will still pass after additional research. This test gives no motivation to “look for bad news” and the research program can be driven by the researcher’s parochial interests.

A more complex test is proposed that takes the “firmness” of risk assessment into account, along with a corresponding research management principle. It extends the familiar "performance allocation" procedure. Second-order risk assessments of individual risk variables are specified as targets. These are set so that if all targets are met, overall risk is judged acceptable, and research effort is directed cost-effectively. As research proceeds and evidence accumulates, the assessment targets are reconfigured and research effort is reallocated. Research is halted when the test of acceptable second order risk is met or is found to be inachievable, and the site is either approved or disapproved.

MANAGING RESEARCH TO EVALUATE IF A RISK IS ACCEPTABLE: APPLIED TO NUCLEAR WASTE SITING

Society requires that a potentially hazardous facility, such as a dam, reactor, pipeline or chemical plant, be established as safe enough to permit its operation. If necessary, data is to be gathered, in a way that properly balances sound and relevant findings, low cost and timeliness. The data gathering can cost much unnecessary time, resources and even public embarrassment, unless there is firm and defensible guidance on how to manage the supporting research. What logic can be used to discipline resource planning in the public interest that is compelling enough to stand up to strenuously defended research, commercial and political interests?

This paper proposes a rationale and paradigm for allocating and reallocating research effort to establish whether the risk from a hazardous facility is acceptable, in situations typified by the nuclear waste siting case. It is based on my experience of working on that problem for eight years, and in particular serving as methodological advisor to two successive heads of DOE's Office of Civilian Radioactive Waste Management (Rusche and Bartlett).

1. THE NUCLEAR WASTE SITING PROBLEM

Nowhere are the issues more critical to resolve than with the disposal of high-level nuclear waste. Although the scale of this problem is exceptionally great, the issues and proposed resolutions typify a wide range of risk management problems, especially where large funds and public controversy are involved.

1.1 The setting

Need for defensible rationale. Nuclear waste disposal is highly controversial and any action is liable to searching political, regulatory and ultimately judicial, review. The National Research Council (1990) and others have questioned the soundness and cost-effectiveness of existing federal approaches to nuclear waste disposal. Logically rigorous principles are needed to assure control of the research management process, in a form that lends itself to systematic review by outside parties. Precious research effort can be wasted, as research contractors pursue their own ambitious research agendas, without clear guidance from their sponsors on when enough research has been done on any particular risk issue or on site risk as a whole.

Acceptable risk. Radioactive releases to the accessible environment must be controlled, in compliance with Environmental Protection Agency standards. In addition, "reasonable assurance" is required that cumulative release of radionuclides over ten thousand years should be below specified limits, with at least 90% probability [1] (EPA, 1985; NRC, 1985). How this is interpreted

---

[1] There are comparable tests for other parameters and other contributing variables, such as groundwater travel time and waste package ainment (not treated separately here), and more conventional regulatory conditions, having to do with hydrogeological features etc. (NRC, 198

determines when enough information has been gathered. It may simply require that the probabilistic test be satisfied on the basis of whatever evidence is available. Certain research activities are required by regulation, such as sinking an exploratory shaft. A more demanding test would be to require that no amount of further research could plausibly shift that finding. There is no general consensus on which test is appropriate.

1.2 State of the art: Performance allocation.

Probabilistic goal allocation is a familiar general strategy for directing research effort selectively at different elements of a system whose performance is to be evaluated, and has been widely used in risk assessment (Apostolakis, 1985; Sung and Cho, 1989; Hunt and Modarres, 1984; OECD, 1989).

DOE's Office of Civilian Radioactive Waste Management has developed a variant, "performance allocation", adapted to the nuclear waste case (DOE, 1988, chapter 8). A hierarchy of performance measures (variables), corresponding to regulatory issues to be resolved, is specified. Each is assigned a target assessment, characterized by a numerical goal and a required "confidence" of its being met. For example, groundwater travel time through a certain geologic stratum has been assigned a goal of 1000 years with "high confidence needed" (DOE, 1988, table 8.3.5.12-1). These targets are set ("allocated") to be consistent with the top level of target assessment: total release less than the regulatory goal with "high confidence". The research strategy is designed to generate data which will support the target assessments.

From among acceptable "performance allocations" one is chosen which appears to be technically achievable and minimizes cost and delay. As knowledge becomes updated, this performance allocation may be revised and with it the corresponding research strategy. Whether research is halted as soon as top level target performance is met depends on the test of acceptable risk.

Appealing as this approach is, I know of no cases where it has been developed and implemented to the point of significantly influencing how resources are actually used. [LIT REVIEW]

1.3 A hypothetical case

To illustrate the problem and possible solutions, I will hypothesize a situation, which closely resembles a real current case. That is the Department of Energy’s project, started in the late 1980s, to establish the suitability of a nuclear waste site at Yucca Mountain in Nevada. Based on a multi-billion dollar, multi-year “site characterization” effort, DOE was to seek a license from NRC to build a high-level waste repository there, or to abandon that site. Numerical and other specifics not relevant to our methodological concerns have been changed in what I will call the “Lacquer Mountain” case.

Site characterization. A team of scientists and engineers have settled on a list of factual questions to be answered, in order to determine if Lacquer Mountain is a suitable site for a repository.

Suitability depends critically on whether significant amounts of radio-nuclides can be excluded from the accessible environment for 10,000 years. Questions include ground-water travel time, radioactive retardation, gaseous release, human intrusion, with various sub-questions.

Research contractors specializing in these issues have been selected, and have submitted estimates of what it would take to "adequately" research these issues. It totals $4 billion over ten years. After the players make their cases, the government apportions a budget of $2 billion, much of it devoted to groundwater travel (which has the strongest research team and has traditionally dominated siting studies), and most of that is to study one of the several geologic strata water must pass through. (Modeling all of them has been judged to be too difficult).

Work begins and understanding of relevant phenomena evolves. Early findings indicate that gaseous release is a more serious source of concern than water-borne radionuclides and much less is known about it. The groundwater team argues against funding being reassigned to the gas team (which has fewer mouths to feed and therefore less clout). The social science team (which is even smaller and carries almost no clout) argue that human intrusion may be the greatest risk and that major research is needed on its massive uncertainties. Lacking any effective analytic weapon with which police the competing factions (including performance allocation as currently developed), the government leaves the budget allocations basically unchanged. Congress cuts the total budget substantially under political pressure, and what is left is fought over by the research contractors and their DOE champions as a conventional exercise in bureaucratic politics.

I will explore two analytic approaches to rationalizing this process, which attempt to refine and elaborate performance allocation. One is based on a well-known simple test of acceptable risk; the other on a more complex test which explicitly incorporates “reasonable assurance”. These developments elaborate Brown (1993, 1988)

2. RESEARCH PLANNING USING A SIMPLE TEST OF ACCEPTABLE RISK

1.1 ***Interpretation of the simple test***.

Risk as a current probability assessment. At any stage in the site characterization research process, a probability distribution on a performance measure, based on available evidence and some consensus of scientific and methodological expertise, can be constructed, as shown by the wider curve in Figure 1. Following the graphic conventions of Brown (1991), this is a "simple assessment" of release; the dashed line is its 80% credible interval and the short vertical "cross hatch" is its mean.

[FIGURE 1]

Target risk assessments. A simple test of regulatory compliance is whether this distribution is acceptable--in this case, whether there is at most 10% probability that the total release exceeds the limit. The narrower curve of Figure 1 shows one illustrative "target assessment" that just meets this test. (It logically has the same expectation as the current assessment.) The current assessment shown says that waste isolation is **not yet** acceptable. Although release is **expected** to be below the goal, probability is less than **90%** that it will be.

***

1.2 Partitioning a simple assessment

A probability distribution for release (whether current or target) can be derived indirectly by partitioning it into a hierarchy of nested functions of more accessible variables, as illustrated schematically in Figure 2. Each variable in the hierarchy is related, in a predictive sense, to variables below it. For example, at the top level, total release can be predicted from three different releases by mode: gaseous, water, and direct human intrusion. Going down the hierarchy: the water borne release can be predicted from transport through geologic and engineered barriers; geologic transport from transport through several geologic strata; each of these from ground water travel time and a radionuclide retardation factor; water travel time from effective rock porosity and hydraulic gradient parameters; and these finally from directly measurable characteristics, such the density of rock core samples and the hydraulic head at a particular location.

[FIGURE 2]

The assessments shown have no special significance, since no scales are shown. However, if the top level corresponds to the current assessment (wider curve) from figure 1, and the three "release mode" distributions below it are on the same scale, their relative spans would indicate that, on current evidence, there is least uncertainty about what water-borne releases would be, and most uncertainty about releases due to human intrusion, with gas-borne releases in-between. Note that some variables, such as those that characterize geologic processes are uncontrollable, and research can only improve their assessment. Other variables, such as engineering performance, are controllable, and can therefore be enhanced by deliberate effort. Assessment of total system risk accordingly needs to consider the effectiveness of such human intervention.

Relation between assessments in the hierarchy Our proposed implementation of hierarchically partitioned assessment conforms to the well-established logic of Bayesian statistics (Raiffa and Schlaifer, 1962). According to statistical theory of random variables (Kendall and Stuart, 1961), if any variable is expressed as an exact function of other variables (arguments), the distribution of that variable is implied by the joint distribution of the arguments. Thus the assessments at any level in the partitioning of release (including the top) can, in principle, be inferred from those below it, *subject to certain critical conditions.* . The mathematical linkage must be an identity, i.e. true by definition. Where the model at one level is only imperfectly *predictive* of the variable above it, an appropriately defined error term must be added. Thus, the variables at any given level in figure 2 are *not* completely determined by the variables below them, only predictable with some imprecision. This is true, no matter how exhaustively the variables are listed (rather than only illustratively, as in Figure 2). There will always be some ambiguity about the appropriate linking model, and always some residual error, even with the best model (as in regression).

That error may be treated as negligible in some cases (for example, releases through gas, water, and human intrusion can be added to derive total release very precisely if any other mode of release is viewed as virtually inconceivable). On the other hand, predicting ground water travel time from a given set of geologic variables and any given mathematical flow model may be quite doubtful. (As witnessed by disagreements between geologists on the relative merits of certain fracture flow and equivalent-porous-media flow models.) Thus the top-level target assessment cannot be routinely inferred from lower level assessments (e.g., by simulation), no matter how ambitious the intervening physical process modeling has been. In particular, the target uncertainty will be seriously understated if no allowance is made for modeling and residual error. Nevertheless, even without error terms, the assessments are *bounded* by those below them, permitting a consistency check on the hierarchy of assessments.

Dependence between assessments. Since it is the *joint* distribution of arguments which is determining, some account needs to be taken of at least first order dependence (higher moments can probably be ignored). Neglect of dependence generally leads to underestimating uncertainty (variance) at higher levels, since correlations are typically positive. Expectations are little affected (and not at all if the partitioning function is linear) and are simply approximated by substituting lower-level expectations in the model. More precise algorithms are available to handle significant dependencies in non-linear partitions.

1.3 Practical implementation of partitioning.

Computation. How the required assessments should be made goes beyond the scope of this paper, which is to propose an analytic framework to accommodate whatever assessments the best available data and inference methods provide. Nevertheless, we do need to show that the necessary computation is feasible. If all input distributions are specified, standard Monte Carlo simulation techniques can be used. Influence diagrams, and well-developed software to implement them (such as DEMOS), provide a convenient format (Oliver and Smith, 1990) and handle interdependencies elegantly. However, they may be unnecessarily costly and time consuming to run, and burdensome to provide assessments for.

Fortunately, simple formulas using means and credible intervals as inputs are often adequate, particularly for common additive and/or multiplicative partitions (for example release is additive by release mode). The variance of a sum is approximated by the sum of the variances (plus covariance terms) and the rel-variance of a product by the sum of the rel-variances (plus relative covariances). Where the component distributions are well-behaved (bell-shaped), this leads to a simple square root formula approximating target mean and credible interval from component means and intervals, in the independent case (Brown et al. 1992). Dependencies between assessments (such as covariance) can be estimated using conditional expectation elicitation, and the simple formulas can be extended to accommodate them.

Judgmental partitioning. Any formally derived assessment can also be attempted using professional judgment. This is not necessarily inferior and usually much cheaper, but it requires the judge to have all relevant considerations in his head, and in this case an ambitious intuitive mastery of statistical decision theory and of the realities of site characterization. Otherwise there needs to be some effective way of linking complementary expertise. This normally requires some kind of formal model with at least as many elements as there are judges.

1.4 Identifying acceptable performance allocations

The above discussion refers to a *current* assessment of performance and its determinants, based on existing knowledge (with assumptions about engineering design). Acceptable risk, on the other hand, refers to target assessment (e.g. 90% probability of meeting EPA release limits). Performance allocation (as DOE calls it) involves picking *target* assessments for the hierarchy of predictive variables which are expected to lead to an acceptable top-level target assessment.

Allocating complete target distributions To test precisely whether lower level targets are consistent with the top-level target would require full distributions for all predictive variables, plus modeling error (and some allowance for dependence, since positive correlations widen the top-level distribution and weaken compliance).

[FIGURE 3]

Figure 3 schematically illustrates a few hypothetical target distributions. (They may not be exactly consistent as drawn and do not take dependence into account.) However, the **current** assessments of figure 2 would have broader distributions throughout, than these **target** assessments, but with the same mean (as illustrated in Figure 1). A target assessment hierarchy is not unique. For example, the distribution for water-borne release could be made looser and the gaseous or human intrusion release could be made tighter, without affecting the distribution for total release. One of the many "acceptable" combinations of target assessments must be picked.

Summarizing target assessments: Goal/confidence. Following DOE practice, we summarize each target distribution by a numerical goal and the probability of meeting it, which we will call the goal/confidence
for the variable. Together with the mean these capture the key properties of the distribution. (Other parameters such as variance needed, say, for propagating uncertainty can be approximated by assuming normality or log-normality). Note that the goal line is at the *upper* end of the distribution if *less* of the variable is better (e.g., release) and conversely (e.g., travel time).

Identifying an acceptable performance allocation It may be reasonable to set the targets so that the top level EPA target (goal/confidence of limit/90%) would *expectationally* be just met. Meeting the top-level target would not be *assured* because of the residual error uncertainties and dependencies. The larger and more positive (respectively) these are assessed, the more stringent the other goal/confidences need to be. At the bottom of Figure 3 for example, if the target assessments for ground-water travel time and radioactive retardation are both met, any modeling error will make the higher-level target assessment for net transport through Stratum B wider. (Dependence is may negligible here, but not for transport through different strata, one level up).

To assure a given confidence at the top-level, progressively less confidence is needed at lower levels (if residual uncertainty is low). For example, if the goals for gaseous, water, and human intrusion releases were each a third of total release goal, each could be met with *less* than 90% probability for total release confidence of 90% to be met. (The standard deviation of a sum is less than the sum of the component standard deviations with independence.) Furthermore, as Figure 3 shows, the goal line does not have to partition off the same area of each distribution at a node: it can be characterized by any arbitrarily chosen fractile, in this case a goal/confidence. If a single "confidence" (probability) is used for all variables, only one number, the goal, needs to be specified. Alternatively, the *goals* can be constrained to be consistent with each other throughout the hierarchy (and the confidences let to vary).

Examples of performance allocation in DOE documents. In DOE's published Site Characterization Plan, illustrative target assessments for performance allocation are given (DOE, 1988; tables 8.3.5.12-1,2). For example, at the "major parameter" level in our hierarchy in figure 2, ground water travel time through stratum A (actually Calico Hills) is assigned a goal of 1000 years with "high confidence needed", defined as "at least two standard deviations below the mean". We interpret this to mean that the target assessment, if approximately normal, gives only about 2.5% probability that the true travel time will fail to meet this goal (bottom-left curve as drawn in Figure 2). Again, the flux parameter q (below those shown in Figure 2), is currently assessed at "low confidence" (.5 prob) of meeting a goal of .5 mm/yr, while the target assessment is "high confidence" (97.5%). However, there is no indication in the Plan that comprehensive sets of target or current assessments had been specified, such that higher-level assessments in the hierarchy could be derived or estimated.

1.5 Setting the appropriate target allocation.

The task of setting targets has two parts: predicting goal/confidence from given lower levels (including modeling error) in order to test that higher level targets are met (an issue in the propagation of uncertainty); and picking one combination of targets (performance allocations) to guide resource allocation among information gathering activities.

Acceptable performance allocations. As noted, many different performance allocations can meet the first test, prompting different information gathering strategies. For example, a severe goal/confidence could be set for ground water travel time through one geological stratum (as in the above example for Calico Hills), and much weaker targets for all other strata through which radionuclides would need to pass. This leads to a research strategy of putting primary reliance on low ground water travel time through that stratum, with relaxed requirements for the others. In the extreme, one would allocate all site characterization resources to gathering information on that one variable and accepting the current state of knowledge on all of the others (e.g., transport through other geologic strata and radionuclide retardation in that stratum). In that case, target and current assessments would coincide for all variables except one.

In the absence of any more systematic procedure (such as the statistical propagation of uncertainty, which may be prohibitively burdensome), trial and error can be used to identify performance allocations which conform to top-level performance--and appear achievable (see next).

<u>Selecting the most appropriate performance allocation</u>. Which of the acceptable performance allocations makes "best" use of available resources and time? One approach is to minimize the expected cost of achieving the top-level performance targets; another is to maximize the probability of regulatory compliance, i.e., achieving a final performance assessment which meets the top-level target. However, these seductively simple formulations may not adequately fit institutional and technical reality. The first may require a more complex objective function including time urgency. Neither acknowledges that targets may be modified as site characterization unfolds (though the process could be repeated periodically.)

Setting probabilistic goals must take account of what it will take in cost to achieve them, and therefore what research tasks are available to achieve them at what cost, and with what impact on top level uncertainty. If competing tasks address the same sub-tree in the hierarchy of variables, one need look no further than impact on the lowest higher order variable they share. One research task may, however, address several different variables. (For example, sinking an exploratory shaft, might cast light on more variables throughout the hierarchy and so deserve high priority.)

All such considerations would be taken into account, when determining goal/confidences throughout the hierarchy. If the linkages within the hierarchy have been modeled, the expectational impact of alternative goal/confidence patterns can be calculated and compared on cost and feasibility. For example, one can track the impact of shifting resources from water to gas-borne release, from Calico Hills to other geologic strata, or from groundwater travel time to radioactive retardation. Formal optimization (say using well-known resource allocation algorithms) may be unmanageably burdensome, but the concepts can give illuminating insight for more informally allocating available dollars among research activities.

The above discussion is in essence a formalization of a qualitatively expressed reasoning developed by the Department of Energy Nuclear Waste Program (OCRWM). It provides an explicit discipline for program management conforming to well-established statistical theory and contains no radical new concepts. However it ignores some issues at the heart of any realistic interpretation of regulatory requirements that have not been generally, or at lest explicitly, recognized.

2 A COMPLEX TEST OF ACCEPTABLE RISK BASED ON PRESCRIBED RESEARCH

2.1 Paradoxical implications of simple test

If "simple compliance" assumed above were sufficient, the preponderance of scientific data already available appears to indicate that this test is already met for the Nevada site. Widely reviewed and largely unchallenged studies (Keeney and Merkhofer, 1987) suggest that the Nevada site comfortably meets not only the isolation (cumulative release) test discussed, but also other performance tests specified in regulation (e.g. ground-water travel time). This would suggest that no further information needs to be gathered--at least beyond the minimal characterization research explicitly required by regulation (e.g. construction of an exploratory shaft and in-situ testing).

Paradoxically, even though characterization can be *expected* to confirm acceptability, the more thorough it is, the more likely it is to yield a major surprise which would overturn that conclusion. Anyone (say a Nevada-based research contractor) with an interest in seeing Yucca Mountain approved would be motivated to gather as *little* relevant information as possible! It might even be argued that no characterization be required by regulation, if simple compliance were demonstrated, though dropping it would surely be politically infeasible. It would also be inconsistent with the NRC requirement of "reasonable assurance" that the simple probabilistic test be met. We interpret that to mean that, in addition to simple compliance, an acceptable risk assessment should be "firm," in the sense of being relatively "unshiftable". Achieving such firmness is presumably the motivation for the characterization requirement in the first place. We call this "complex compliance". DOE implicitly recognizes this requirement by giving high priority to tests which would uncover "potentially disqualifying conditions". We seek an explicit and defensible rationale that can be turned into an operational resource allocation principle.

2.2 Theoretical basis of approach

Assessment uncertainty. The concept of second order assessment uncertainty (Brown 1991a?) shown in Figure 4. Suppose that one of the regulatory issues to be resolved deals with cumulative release and that, for a given site, Figure 4a (the same as Figure 1) represents the current "simple assessment,"[2] i.e., a probability distribution on total release, based on some consensus on present information, before site characterization. The EPA 90% goal, L, which must be met with at least 90% probability is indicated by a vertical line. (The dashed bar and cross-hatch below the figure represents an 80% credible interval and a mean.) The 90th percentile is comfortably below that goal; so simple regulatory compliance on this issue is assured (if NRC accepts the distribution).

[FIGURE 4]

Predicted-assessment uncertainty. Simple assessments may shift over time, giving rise to present "assessment uncertainty." If the shift will be after some specified development, we call it a "predicted-assessment uncertainty" (PAU); for example after the completion of a specific characterization program. Two possible shifts in the simple assessment of release in figure 4a after such research are shown in Figure 4b. One meets the 90% goal, as a result of a favorable characterization and the other does not. Suppose our predicted assessment uncertainty is such that we give 10% chance that characterization will prove to be more and less favorable than these, respectively. Figure 4c represents a complete PAU distribution, but of a special kind. It predicts where the critical 90th percentile will fall, rather than, say, the mean as in Brown (1991a). Figure 4c shows 85% probability that the new 90th percentile will be below the goal, i.e., a 15% chance of failing to meet the L/90% goal/confidence and therefore not complying with regulation. We use a thick solid bar to represent the 80% credible interval of a PAU.

2.3 Licensing based on "adequate characterization"

The regulatory and program management significance of this hypothetical assessment uncertainty depends on one's view of NRC's licensing position. NRC (1985) regulation requires that certain site characterization activities, such as sinking an exploratory shaft, precede license application. One plausible regulatory principle would be for NRC to grant the license if the ensuing simple assessment meets the 90% test.

Not-yet-in-compliance case. The tightening of the simple assessment curve in figure 1 shows how non-compliance might turn into compliance. More generally, compliance is most likely the smaller the tail of the PAU distribution is to the right of the goal line in Figure 4c. Clearly, the more relevant research that is done, the tighter one can expect the simple distribution to become around the current mean which is shown in the acceptable range. (We are only considering the case

---

[2] Some terms introduced in Brown (1991a) have been changed. "Simple assessment" was called ctual assessment" and "predicted-assessment" was called "post-assessment".

where the mean is already below the goal, which is realistic since presumably the site would not otherwise have been put forward). If (as in the wider curve of figure 1) it does **not** currently meet the simple acceptable risk test then the more research is done, the greater the chance that it will comply afterwards (as in the narrower figure 1 curve).

Already-in-compliance case. However, in the case shown in Figure 4a, the simple test is already met. One *expects* research to give a narrower distribution with less of the tail beyond the goal, thus meeting the test even more convincingly. But there is the *possibility* of producing a failing assessment, i.e. a curve with its tail beyond the goal, and the probability gets larger with more research. (The curve in figure 4c flattens with more area beyond the goal and thus more probability of non-compliance). Conversely, doing no research guarantees remaining in compliance. Therefore a site supporter would be motivated to do the least diagnostic research that NRC regulation allows! In section 4 we will consider an alternative licensing principle which addresses this paradox.

We will now consider the first case, and the research strategy most likely to maximize the probability of post-research compliance, if simple compliance is *not* achieved currently.

2.4 Deriving top-level PAU

Propagation of assessment uncertainty. The top level assessment uncertainty called for in figure 4c can be derived by propagating lower levels of assessment uncertainty, similar to simple assessment. Figure 5, with the same structure as Figure 2, adds hypothetical predicted-assessment uncertainties, which predict simple assessments that will be made after an "adequate" characterization. (Heavy bars represent credible intervals on simple 90% fractiles, and vertical lines correspond to 90% goals as in Figures 3 and 4.)

[FIGURE 5]

The top level assessments correspond to Figure 4; i.e. the simple assessment for total release is well below the goal, as indicated by a dashed line to the left of the goal line below it. After research, however, the goal is quite likely *not* to be met, as indicated by the solid bar for predicted-assessment uncertainty extending to the right of the goal line (as in Figure 4). An objective of site characterization planning might be to maximize the probability that this goal will be met, i.e., to make this top-level PAU bar as far to the left as possible.

This depends on the pattern of lower level PAU, which may differ markedly from the simple assessments, because variables with large simple uncertainties do not necessarily have large PAU uncertainties and vice-versa. Whether a simple or complex test of acceptable risk is used thus has significant implications for resource management.

Interpretation of illustrative assessments. For example, at the first "release mode" level in figure 5, simple assessment uncertainty and differ distinctively for each mode. Water-borne release is currently the most thoroughly researched and has the least simple uncertainty, and so has the shortest dashed line. Characterization may shift it, but modestly, with a fairly short heavy bar. Both lines are to the left of the 90% goal line; not only is water-borne release expected now to meet the goal with over 90% probability, but after characterization research it can be confidently predicted to still meet it. On the other hand, human intrusion (including any political, social, and institutional disruptions over ten thousand years) make it simple assessment highly uncertain (long dashed line). However, what little can be done to narrow that uncertainty has already been done (e.g., appraisal of minable resources) and PAU accordingly has a very tight distribution (short heavy line). Gas-borne release is a different case again. Based on more limited evidence, it has more simple uncertainty than water (longer dashed line, but less than human intrusion). It appears to meet the 90% goal (dashed line to left of goal); but research may well shift it into the unacceptable region (longest heavy bar, crossing goal line).

Implications for research strategy. Figure 5 corresponds to a specific research plan (allocation of resources). The longer the PAU bar at the top level, the more diagnostic that plan is expected to be. Alternative plans can be evaluated in terms of their impact on this bar.

The research on gas-borne release is judged here to have more impact than research on either waterborne or human intrusion releases, even though it is between them on simple uncertainty. This implies we should move resources from gas to human intrusion if the resulting shortening of the gas PAU bar would more than compensate for lengthening the intrusion bar (in terms of impact PAU one level up). If the proposed human intrusion research already covers virtually everything that can usefully be discovered, and taking resources from gas would mean cutting out productive research, the answer is presumably "no." The practical implication is that more effort be devoted to gas-borne release (and less to human intrusion).

At the "sub-barrier performance" level, strata A and B show comparable simple uncertainty (similar dashed lines) but A has more PAU (longer heavy bar). If the greater shift is because more research has been planned for A and shifting resources to stratum B would permit a relatively large lengthening of its PAU, it would be worth doing. However, if A's longer PAU bar is because the same effort yields more information than for B, the proposed balance may be appropriate. (Note that how changes in low level PAU translate into upper level changes depends on how the variables combine. For example, PAU for release modes, being in parallel, are handled differently mathematically from geological strata which are in series).

3 A MORE COMPLEX TEST OF ACCEPTABILITY BASED ON "FIRMNESS" OF SIMPLE ASSESSMENT

WE noted earlier that this argument runs into difficulties if simple risk assessment is *already* acceptable, since increasing top level PAU (lengthening heavy bar) due to diagnostic research now *reduces* the probability of simple compliance after research. Suppose now, in contrast to the "prescribed characterization" assumption above, that acceptable risk depends on the "firmness" of the risk assessment; i.e., NRC will only license if it accepts, not only that the simple risk assessment meets the 90% goal, but also *any* further characterization is unlikely to overturn the simple compliance test. This may, in fact, be what is meant by the regulatory requirement of "reasonable assurance" (NRC, 1985).

3.1 Ideal-assessment uncertainty.

We define ideal-assessment uncertainty (IAU) as a distribution on where the simple assessment might shift to, given *unlimited* (but technically feasible) resources for information gathering and analysis, i.e., an ideal characterization program (Brown, forthcoming). Ideal-assessment uncertainty is always greater than predicted-assessment uncertainty, because IAU represents more data and therefore more potential for shift. It can be used as a useful guide to where application of research resources is likely to be most productive. Thus, if IAU for intrusion is barely greater than the already tiny PAU, it means there is nothing much more that can be done to reduce uncertainty here, and so little to be gained by devoting more resources to it. Conversely, if IAU for gas-borne release is much greater than its PAU, there is much scope for research additional to that planned.

Propagation is exactly parallel that for PAU and can be represented similarly (with a line instead of a bar representing 80% credible interval).

3.2 Reformulation of acceptable risk test.

Figure 4c could be recast for IAU, with research now being interpreted as unlimited. If, as shown, it did, in fact, represent ideal-assessment uncertainty, it would mean that, although the simple compliance test is met currently, there is a 15% chance that given enough characterization it would fail. 15% may be unacceptably high, and NRC may, on those grounds, decide the site is not yet licensable. Opponents of this waste site do in fact argue along these lines.

[FIGURE 6]

Regulatory implications. Figure 6 shows various uncertainties relating to a specific research plan. Under "current assessment", the top two rows address the same issues as figures 4 and 5: i.e. what does the risk of radioactive release look like now and what is it likely to look like after planned research (paying specific attention to the .9 fractile). As before, the current simple assessment shows that release meets the EPA 90% release goal, but less than 90% that it will still clear it after planned research. The third row indicates a greater probability (say 20%), given **ideal** research, that the simple .9 fractile will shift enough exceed the release limit.

Suppose the regulatory requirement is that the probability of the ideal .9 fractile exceeding the limit should itself not exceed 10%. (It could be 50%, since existing regulation is silent on this issue.) Now an objective of research might be, not only to end up with a simple assessment whose .9 fractile clears the limit (which the first row of figure 6b shows it doing easily); but to also with the ideal .9 fractile clearing the limit (which the bottom row shows it just doing).

Note that the two assessments in figure 6b are just targets: examples of assessment that might emerge after research--and if they did they would prove acceptable (under our current assumptions). Predicting what they will actually be after research involves higher order probability assessment (third order in the ideal case, which is already second-order). Figure 6b could be taken as modally predicted or expected distributions. We noted earlier the paradox that the more effort devoted to researching any variable, such as release, the greater the chance that the simple assessment will shift into the unacceptable zone. However the **expectation** is that it will improve (as with the first illustrative assessment in figure 6b). More important, the ideal-assessment is also likely to shift, this time from unacceptable to acceptable, as shown at the bottom of figure 6b.

The guiding principle now for planning actual research would presumably be to maximize the chance of achieving such an acceptable ideal-assessment uncertainty, remaining **after** the actual research. Attempting to rationalize this formally appears to involves **third-order** probability concepts (i.e., the probability of a second-order assessment). We do not propose that as cognitively practical even for statistical specialists. However, we do think it is realistic for research planners to bear these considerations in mind, as they assign resources to research activities, and reassign them as results unfold. In fact they have no choice if the test of regulatory compliance includes this kind of consideration.

3.3 <u>Goal allocation for complex compliance</u>

The principles for probabilistic goal allocation in the simple compliance case (section 2), can be adapted to guide research management in either of the complex compliance cases. The resource allocation and reallocation procedure DOE adopts will now be the formal equivalent of that proposed before, in Section 2, with assessment uncertainties (and their goal/confidences) playing the role that earlier was played by simple assessments. Again determination of the targets requires information on cost effectiveness of research on each variable, which relate reduction in the width of the assessment uncertainty bar (i.e., the probable shift in conclusions) to the application of increasing resources. In the "ultra" complex case, the guiding principle is to allocate available research resources so that, after research, the resulting ideal-assessment uncertainty bars (including those for modeling error) produce an acceptable top-level IAU bar (which will also meet the simple compliance test).

4 CONCLUSIONS

4.1 <u>Findings</u>

We have proposed what we believe is a tenable rationale for determining acceptable risk, and d a strategy for managing the research necessary to establish it. It is predicated on two possible interpretations of regulatory requirement that go beyond the simple first order probability tests of compliance currently advocated in EPA standards and elsewhere. The "ultra" complex variant is probably the more promising as a regulatory principle.

It builds on a treatment of second-order assessment uncertainty, proposed in Brown (1991a). This rationale has been illustrated in the context of verifying the suitability of a federal nuclear waste repository. In a highly contentious case like that, it may prove of great practical importance to be able to demonstrate (to courts or regulators) that the acceptability of a proposed repository and the defensibility of the research strategy that led up to it can be soundly validated.

It may not be essential to show that the rationale itself was actually used every step of the way in making the research planning decisions, provided it can be shown, after the fact, to have conformed to the rationale. However, it may, in fact, be critical to apply the rationale explicitly (if not necessarily in detail) from time to time to assure proper management of scarce resources. Without such discipline the private (e.g. scientific) priorities of the research teams may take precedence over the judicious balancing of social protection and cost.

4.2 Practical Implications

Application of any of the logic proposed here may have significant impact on currently proposed strategies. For example, it may demonstrate that spreading resources evenly over several geologic strata representing serial barriers is more cost effective than devoting all resources to one primary barrier. An operational heuristic for achieving (or testing for) optimal allocation would be to allocate available dollars such that the marginal "yield" for all activities is equal. This is a well known principle of economic optimization, for cases where "yield" is some simple quantity such as "profit." In this case "yield" would have to be something like "incremental probability of regulatory compliance."

It may not be necessary for the practical implementation of these principles be very detailed. Any quantitative or other explicit procedures, specified in advance, can only approximate the considerations that will emerge as relevant at the time decisions are actually made. Therefore these procedures should not be considered as binding on the decision process, only as an illuminating aid and an important element in validating decisions after the fact.

More specifically, we are not suggesting that the principles proposed here can be readily translated into an operational methodology to mechanically produce recommended allocations of goals of resources. Rather it provides for a test of the appropriateness of allocations currently adopted or under consideration. These may be based on professional judgment, which more informally takes our logical concepts into account. It also provides a convenient framework within which to present competing arguments about what should be (or should have been) done. The approach can be used to validate what ever research management strategy is proposed (even if it was not used in the first place to derive the strategy). It should help the risk manager, DOE in this case, to withstand, not only regulatory and legal challenge, but also the resistance of researchers to having their projects terminated for programmatic (rather than scientific) reasons.

4.3 Caveat: Discovering unacceptable risk as a research objective

We have only considered the establishment of *acceptable* risk aa a research objective. Discovering that risk is *unacceptable* is no less important.

Relation to traditional acceptance testing. In conventional acceptance testing theory (Fisher, 1935), candidate tests of whether to "accept" a hypothesis as true are evaluated with a view to minimizing two probabilities: "false positive", i.e. the hypothesis is accepted, though false; and "false negative", i.e. the hypothesis is rejected, though true. The more exacting the test, the better it will fare on the first and the worse on the second. In this case, the hypothesis is that the site meets regulatory requirements. We are interpreting this here to mean that ideal-assessment would show at least 90% probability that release will be less than the EPA limit; corresponding to and IAU whose .9 fractile is at the goal. Acceptance means DOE concludes it has enough evidence to apply for an NRC license.

The implicit acceptance rule presented above directly uses the probability of false positive. Figure 6 reflects the case where that acceptance probability is set at a low 10% (bottom of fig 6b). Conversely, the corresponding false negative probability would be relatively high, but its exact value would need additional assessment. (It would depend partly on the prior probability of the hypothesis reflected in the bottom of figure 6a.) The appropriate trade-off between the two types of error probability would be a matter for regulatory policy.

4.4 Further Research Needed

To turn a logical rationale into an operational methodology and extending it to include the false negative/positive issue will require much prescriptive decision research. What input can be cognitively supplied, what output can be institutionally used and what logical algorithms are appropriate to link the two? Highly skilled technical work is needed to develop the specific procedures and software to implement the rationale in a practical context. Although we have taken as our illustration a problem of unusual practical importance, the issues addressed arise in virtually any case where, for regulatory or other purposes, the acceptability of a risk needs to be determined. "Acceptable risk" may need to be conditioned on circumstance, to take account of trade-offs between risk and other considerations. If there is an issue of what research to do when enough has been done, our approach should in principle be applicable. This is true, whether we are talking about genetic engineering, industrial safety, or the health risks of consumer products.

ACKNOWLEDGEMENTS

The author is grateful to Jake Ulvila, Lee Merkhofer, Norm Eisenberg and two anonymous reviewers for very helpful comments and suggestions. The work was supported in part by the National Science Foundation, Decision, Risk and Management Science Program.

REFERENCES

Apostolakis, G. *Some issues related to goal allocation and performance criteria,"* Paper M2 4/3, 8th International Conference on Structural Mechanics in Reactor Technology, Brussels, Belgium, August 19-23, 1985.

Brown, R.V. *Research and the credibility of estimates.* Boston, Ma., Harvard University, Graduate School of Business Administration, Division of Research, l969.

Brown, R.V. *Proceedings 1988 Annual Meeting Decision Sciences Institute* November 21-23, 1988. Las Vegas, Nev., Characterizing a Nuclear Waste Site: A Logical Framework for Allocating Resources, 1988.

Brown, R.V. Assessment uncertainty technology for making and defending risky decisions. *Journal of Behavioral Decision Making*, 1991a.

Brown R.V. Impersonal Probability: a viable and useful construct? *Journal of Risk and Uncertainty*. 19[xx].

Brown, R.V., Lilien, G., Ulvila, J. New Methods for Estimating Business Markets, Journal of Business to Business Marketing, 1992.

Brown RV. Managing research to establish acceptable risk: probabilistic goal allocation with application to nuclear waste siting. Dept. of systems Engineering. George Mason U. Working paper WP930302. 1993.

Covello, Vincent T. and Merkhofer, Miley W, *Risk Assessment: Methods and Approaches for Quantifying Health and Environmental Risks*, N.Y., Plenum Press, in press.

Diamond, William J. *Practical Experimental Designs for Engineers and Scientists,* 2nd Edition, N.Y. Van Nostrand Reinhold, 1989.

Hunt, R.N.M. and Modarres, M. Integrated economic risk management in a nuclear power plant. In *Proceedings of the Annual Meeting of the Society for Risk Analysis*, October 1984, Knoxville, TN, Plenum Press, New York.

Keeney, R.L. and Merkhofer, M. A multiattribute utility analysis of alternative sites for the disposal of nuclear waste. *Risk Analysis*, *2*(7), 1987.

Kendall, M.G. and Stuart, A. *The advanced theory of Statistics.* London: Hafner, 1961.

National Research Council (Board on Radioactive Waste Management). *Rethinking high-level radioactive waste disposal*. Washington, D.C.: National Academy Press, 1990.

National Energy Association. Proceedings of an NEA Workshop on *Uncertainty Analysis for Performance Assessments of Radioactive Waste Disposal Systems*, Seattle, February 24-26, 1989, OECD, Paris.

Oliver, R.M. and Smith, J.Q. (Eds.) *Influence diagrams, belief nets, and decision analysis*. NY: Wiley, 1990.

Sung, S.K. and Cho, N.Z. Determination of performance criteria at hierarchical levels in a nuclear power plant, *Reliability Engineering and System Safety*, 24:231-256, 1989.

Department of Energy. *Issues hierarchy for a mined geologic disposal system.* Office of Geological Repositories, U.S. Department of Energy, August, 1987.

Department of Energy. *Site Characterization Plan: Yucca Mountain Site.* DOE/RW-0199. USDOE Office of Civilian Radioactive Waste Management. Dec. 1988.

? Department of Energy. *A multiattribute utility analysis of sites nominated for characterization for the first radioactive waste repository/A decision aiding methodology*. Washington, D.C.: Office of Civilian Radioactive Waste Management, DOE\RW-0074, 1986.]

Environmental Protection Agency, 40 CFR 191. *Environmental standards for management and disposal of spent nuclear fuel, high-level and transuranic radioactive wastes,* 1985.

Nuclear Regulatory Commission. *Code of Federal Regulations,* Volume 10 Part 60. Washington D.C.: Government Printing Office, 1985.

LIST OF FIGURES

Figure