ISSUES IN FORMULATING PERFORMANCE-BASED APPROACHES TO REGULATORY OVERSIGHT OF NUCLEAR POWER PLANTS

R. W. YOUNGBLOOD* and I. S. KIM
Information Systems Laboratories
11140 Rockville Pike, Suite 500
Rockville, MD 20852 USA
*Corresponding author. E-mail : ryoungblood@islinc.com, ikim@islinc.com

Received May 30, 2005

In recent decades, significant effort has led to risk-informed improvements to regulation. Performance-based approaches also promise significant gains in efficiency (level of safety versus effort). However, significant work remains to be done before performance-based approaches realize their full potential in regulation of nuclear power plants. This paper reviews key concepts related to performance-based regulation, discusses some applications of performance-based approaches, and identifies issues that still need to be addressed. Realistic, experience-based models of licensee performance are still lacking; this makes it difficult to assess the prospective effectiveness of any given regulatory approach, in light of the performance issues that it will actually face. Also, while "compliance" is an intuitively straightforward concept to apply within a prescriptive implementation, its analog in a performance-based approach remains unclear. An overarching theme of the paper is that formal methods of decision analysis are very helpful in developing appropriate regulatory approaches, especially performance-based ones; this theme is illustrated at several points.

KEYWORDS : Performance-Based Regulation, Decision Analysis, Bayesian Performance Assessment, Performance Indicators, Probabilistic Risk Assessment, Risk-Informed Regulation

1. INTRODUCTION

1.1 Purpose

In recent decades, significant effort has led to risk-informed improvements to regulation. Performance-based approaches offer significant gains in efficiency. However, significant work remains to be done before performance-based approaches realize their full potential in regulation of nuclear power plants. This paper reviews the concepts and discusses some applications of performance-based approaches, and identifies issues that still need to be addressed. What shortcomings of current regulatory approaches might be addressed by risk-informed, performance-based regulation? Much has been written on this subject, and it is not the purpose of the present paper to review this discussion in detail. It is widely agreed that the burdens imposed by some current regulatory requirements are not aligned optimally with respect to real safety priorities. Depending on the particular issue, either too much is being spent to achieve a given safety outcome, or safety expenditures are focused in the wrong areas, or perhaps both. Risk-informed and performance-based considerations address both of these conditions.

The approach to regulation varies significantly from one country to another and from one industry to another. Regulation of a given industry can be very different in different countries, even given generally similar safety intent. For example, consider how different countries approach cost-benefit analysis. Some of the examples discussed in the present paper have been influenced by US regulation of its commercial nuclear power plants (NPPs). However, it is not the purpose of this paper to critique the US approach, or to comment on its advantages or disadvantages compared to other countries' approaches. Where specific examples are discussed, the purpose is to illustrate purely technical considerations in formulating an optimal regulatory approach.

As discussed later, performance-based approaches may be applied, whether or not risk information is used directly to set safety priorities. However, there is a very natural relationship between risk-informed and performance-based regulation: performance-based regulation requires that performance goals be set, and using risk models is a very natural way to do this. Accordingly, the remainder of this section summarizes key aspects of risk-informed approaches at a high level.

1.2 Evolution of Regulatory Applications of Probabilistic Risk Analysis (PRA) in the US

Current regulatory requirements on commercial nuclear power plants in the US have been derived primarily from deterministic considerations. For example, two major areas that govern safety system design are: (1) safety system design requirements, largely associated with the General Design Criteria of 10 CFR Part 50 Appendix A [1], expanded upon in Regulatory Guide 1.70 [2] and the Standard Review Plan [3]; and (2) design basis accident analysis guidance of Chapter 15 of Regulatory Guide 1.70, of the Standard Review Plan, and of 10 CFR Part 50.46 and Appendix K, directed towards demonstration of adequate design margins based upon defined acceptance criteria. These requirements lead to the imposition of special treatment requirements on key systems, structures, and components (SSCs) relied upon to satisfy requirements and/or mitigate postulated challenges to plant safety functions. In addition to special treatment requirements, numerous other prescriptive requirements are imposed on these SSCs in the areas of testing, inspection, and technical specifications governing operations. Compliance with regulatory requirements is deemed to provide reasonable assurance of adequate protection.

In the mid-1970's, the Reactor Safety Study [4] offered a complementary perspective on plant safety. It showed that the alignment of safety resources with actual risk significance that resulted from the deterministic approach was not optimal. The emphasis on physically challenging but very unlikely design basis events (e.g., large loss of coolant accident (LOCA)) drove core damage frequency (CDF) from those events to relatively low levels; but mitigating system reliability for more frequent challenges (e.g., loss of offsite power) was not (at the time) commensurate with the frequencies of those challenges. Therefore, beginning in the late 1970's and early 1980's, several regulatory initiatives (the anticipated transient without scram (ATWS) Rule [5], the station blackout (SBO) Rule [6], and so-called TMI (Three Mile Island) requirements [7] such as the auxiliary feedwater system (AFWS) requirement) supplemented earlier requirements on frequently-challenged systems. In addition, PRA began to be applied more systematically to rationalize testing intervals on some equipment, and allowed outage times in technical specifications [8-11].

In 1988, the USNRC mandated the Individual Plant Examination (IPE) program [12]. This program required licensees to examine their plants for vulnerabilities. Full Probabilistic Risk Analyses (PRAs) were not necessarily done in response to this program requirement, but most licensees systematically applied key elements of PRA methodology to search for vulnerabilities at their plants. Plant modifications or enhancements to training and procedures were undertaken as a result of this improved understanding. Some licensees have chosen to build on this investment in analysis, and make increasing use of their analyses in justifying changes to their licensing bases.

In 1995, the USNRC promulgated its PRA policy statement, including the following:

> "The use of PRA technology should be increased in all regulatory matters to the extent supported by the state of the art in PRA methods and data, and in a manner that complements the NRC's deterministic approach and supports the NRC's traditional defense-in-depth philosophy."

Currently, PRA is recognized as a key tool in decision-making related to changes to the current licensing bases of existing plants [13-17], and is required in applications for design certification.

Changes have now been made to the USNRC Reactor Oversight Process (ROP), which uses risk information to guide inspections and assess the significance of inspection findings. Regulatory activities to make changes to regulatory requirements are also proceeding. These include combustible gas control requirements in 10CFR Part 50.44, the new 10CFR Part 50.69 Special Treatment Regulation, and potential changes to the design basis LOCA requirements in 10CFR Part 50.46 [1].

Although risk analysis plays an important role in decision-making for commercial nuclear plants, USNRC decision-making is currently intended to be "risk-informed" rather than "risk-based." The attributes of "risk-informed" decision-making are given in Regulatory Guide RG 1.174 [13]. Risk-informed decisions consider information in addition to the output of current risk models. "Risk-informed" is a concept that applies to regulatory decision processes. However, performance indicators that derive entirely from current PRA treatments can reasonably be called "risk-based." Thus, "risk-based" performance indicators can be applied within a "risk-informed" regulatory decision-making process. Examples will be discussed later.

2. ELEMENTS OF REGULATORY APPROACHES

2.1 Allocation: Deciding "What's Important"

Different approaches to regulation of nuclear power plant operation are shown in Figure 1, which evolved from a diagram appearing in an Industry White Paper [18].

In formulation of a regulatory approach, safety priorities must first be established. For a given technology, certain levels of reliability are needed in systems that perform certain safety functions. This is an example of what is meant by the heading "What's Important" appearing on the left of Figure 1. Under this heading are different possible ways to go about establishing "what's important." One way is frequently called the "deterministic approach," by which is meant the process applied traditionally in the

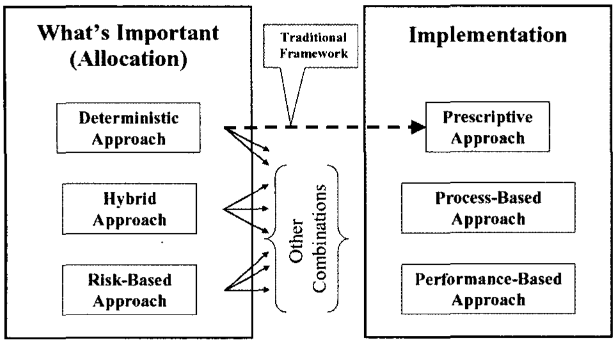

Fig. 1. Regulatory Approaches

US in NPP licensing. This approach includes specification of stringent challenges (including so-called "design basis" events) that are to be mitigated. Demonstration of design capability is to be carried out subject to specified initial conditions and postulated failures, and must be analyzed using specified evaluation methodologies. For example, within a deterministic approach, one might decide that the emergency core cooling system function is important, and needs to be sufficiently redundant to tolerate a single active failure, conditional on a specific initiating event, concurrent with a loss of offsite power.

Within a risk-informed approach, one might decide that post-trip removal of decay heat should have a functional unreliability on the order of 10^{-4} or less. Both of these decisions represent decisions to invest in particular safety features in order to satisfy a higher-level safety objective of some kind: perhaps an implicit standard of protection, or perhaps quantitative safety objectives. It is convenient to have a general term for this kind of decision. Here, and in other documents (Appendix B of NUREG/BR-0303 [19]), the term "allocation" is used for this kind of decision. Note that this usage is intended to be general, and is not restricted to the specific case of assigning quantitative reliability targets.

In some ways, the regulatory approach in the US is implemented as if the design basis events are actually the primary operational safety concerns, rather than simply being used to assure robust system performance and high margin by bounding the physical challenges posed by credible initiating events. This is one reason for the perceived sub-optimality of the current approach: it is sub-optimal to manage the plant as if the most immediate threat to public safety is actually a double-ended guillotine rupture of the largest pipe, concurrent with a loss of offsite power, as opposed to an initiating event occurring thousands of times more frequently. A different basis for allocating is to use a risk model. This has the key advantage of explicitly accounting for the frequency with which certain functions are challenged. In NPP regulation in the US, risk modeling has been used on occasion to inform enhanced requirements on frequently-challenged systems, and to allow greater flexibility in treatment of systems challenged less frequently (or systems having redundant alternate success paths).

Making essential use of risk analysis in safety decision-making imposes a large burden on the risk model. This point has been discussed extensively in recent decades (frequently called "PRA quality"), but is not a focus of the present paper. For present purposes, it is important to distinguish the process of allocation from the process of implementation (discussed below); the major focus of the paper is on implementation. In real applications, allocation and implementation need to be considered together, because the net benefit of a given regulatory approach depends on both, but it is useful not to confuse the two aspects.

2.2 Traditional Approach to Implementation

Determining what performance is needed (i.e., allocation) is only part of the story: it remains to achieve that performance in reality. "Implementation" refers to the task of actualizing the above-described allocation: namely, achieving the performance targets, or making the allocation "come true." In the US, implementation has been approached for NPPs by classifying certain systems, structures, and components as "safety-class," based on the role of these SSCs in mitigating design basis events and satisfying other design criteria, and imposing prescriptive requirements on these SSCs.

The prescriptive requirements include quality assurance (QA), in-service testing (IST), in-service inspection (ISI), and numerous consensus engineering standard requirements aimed at assuring highly reliable performance. These requirements are prescriptive in that they prescribe what licensees should do (e.g., what to test, how often to test, etc.). The requirements are sufficiently specific that compliance can be assessed. Reliability is explicitly mentioned only occasionally in the regulations; compliance with prescriptive requirements is presumed to lead to performance that is sufficiently reliable to satisfy the agency objectives.

The traditional framework consisting of deterministic allocation and prescriptive implementation is indicated by the dotted arrow in Figure 1.

A completely prescriptive implementation is not necessarily optimal. Knowing only that a licensee is compliant with prescriptive requirements, one does not know whether safety objectives are being met: it is possible to be compliant but unsafe. Moreover, licensees are not allowed flexibility in how safety is achieved, and as a result, any inefficiency built into the prescriptive requirements cannot be overcome by a compliant licensee. It is also possible to be safe, while not being compliant.

In general, it is challenging to develop an implementation that is optimal, or at least tries to address issues of burden, regulatory effectiveness, and overall net benefit. This task can be approached using the high-level guidelines for performance-based regulation.

2.3 High-Level Guidelines for Performance-Based Regulation

As given in an NRC White Paper [20], the definition of "performance-based" includes the following elements:

(1) Measurable (or calculable) parameters (i.e., direct measurement of the physical parameter of interest or of related parameters that can be used to calculate the parameter of interest) exist to monitor the system, including both facility and licensee performance;

(2) Objective criteria to assess performance are established based on risk insights, deterministic analyses and/or performance history;

(3) Licensees have flexibility to determine how to meet the established performance criteria in ways that will encourage and reward improved outcomes; and

(4) A framework exists in which the failure to meet a performance criterion, while undesirable, will not in and of itself constitute or result in an immediate safety concern.

The first three of these elements appeared originally in the NEI white paper [18]. Together, they describe an approach that is outcome-oriented and therefore performance-based. For example, suppose that "what's important" is "high reliability" of a particular system. A prescriptive approach may try to achieve high reliability through mandated procurement practices and testing, and regulatory oversight within such an approach would simply assess compliance with these requirements. In contrast to this, a performance-based approach might try to monitor reliability, and intervene only if reliability performance became unsatisfactory. "Licensee flexibility" means that in such a case, licensees would be free to choose the treatment and testing practices (and perhaps even the redundancy) needed to meet reliability criteria.

The fourth element in the definition of "performance-based" was added later by USNRC staff [19, 20]. The first three elements do not explicitly preclude a situation in which the performance criterion is sufficiently relaxed that degraded performance could lead to an accident before intervention was triggered by performance monitoring. The fourth element establishes the principle that the framework must support timely intervention by regulatory staff when performance has declined significantly.

These four elements were later incorporated into the "high-level guidelines for performance-based regulation" as the "viability guidelines" [19]. These high-level guidelines comprise considerations of diverse types that bear on development of new regulatory alternatives. Although called "guidelines for performance-based regulation," the guidelines are formulated to be applicable to proposed regulatory alternatives in general (not just "performance-based" ones). Other guidelines included in [19] are the "assessment guidelines" and the "guidelines for consistency with regulatory principles." The assessment guidelines address whether a proposed alternative achieves the following:

- Maintains safety;
- Increases public confidence;
- Increases effectiveness, efficiency, and realism;
- Reduces unnecessary regulatory burden; and
- Results in a net benefit.[1]

Additional assessment guidelines address the ability of the proposal to be incorporated into the regulatory framework, and the ability to accommodate new technology. This evaluation is to be based on an integrated assessment of the individual guidelines within this grouping. The guidelines for consistency with regulatory principles address whether a proposed alternative is consistent and coherent with other overriding goals, principles, and approaches in the NRC's regulatory process.

An illustrative conceptual example of performance-based regulation is the oversight of occupational dose in areas where the dose rate is not too high. Given that dosimetry is reliable and that results are reported in a timely fashion, the regulator can be reasonably sure whether safety practices are accomplishing the intended objectives (whether adverse consequences are occurring). In low-dose-rate areas, invasive oversight of work practices at a detailed level would be unwarranted. However, at a facility using high-dose-rate sources, a different conclusion would be warranted. If defective equipment or human error can cause immediate fatalities, then dosimetry alone cannot provide the diagnostic information needed to trigger a timely intervention to reverse declining performance trends. This idea is discussed at length in NUREG/CR-6642 [21].

Even for the low-dose-rate situation, the regulator needs reliable information if the regulatory intent is to be met. In this case, prescriptive requirements may be imposed not on work practices, but rather on dosimetry, record-keeping, and reporting. Regulatory inspection could accordingly focus on these areas.

3. FORMULATING REGULATORY APPROACHES

3.1 General Aspects

An implicit assumption in most regulatory approaches is that licensee performance affects safety system performance. One way to think about this is illustrated in Figure 2. When performance is "good," risk model parameters have values that collectively correspond to a satisfactorily low level of risk; when performance is "degraded," some risk model parameters assume values

---

[1] This list was originally based on elements of the agency's Strategic Plan. The Strategic Plan has changed since NUREG/BR-0303 was published. Presumably, the assessment guidelines should track the agency's current Strategic Plan.

for which the level of risk is increased, perhaps significantly. This is seldom treated explicitly in PRA.

For simplicity, Figure 2 is drawn to suggest the existence of discrete performance states. This may be an oversimplification: there may be more than two states, or in some cases, it may be inappropriate to view the states as highly discrete. In general, however, the idea that "good" and "degraded" performance exists is quite widespread. From that point of view, it is almost surprising to consider that many attempts to extract reliability parameters from operating experience formulate their results in terms of measures of central tendency (mean, median) of a unimodal distribution, albeit perhaps a broad one. This point will be taken up later.

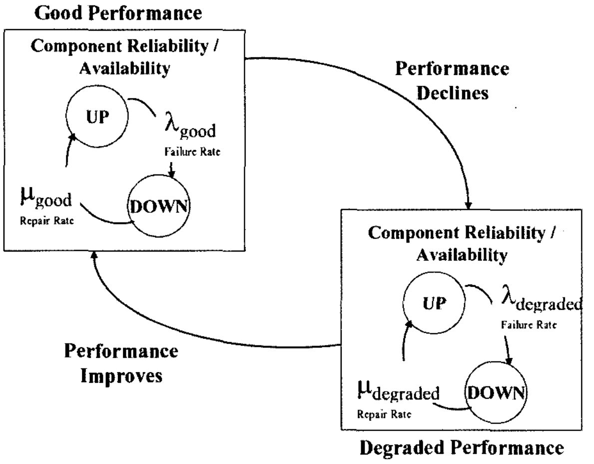

Fig. 2. Effect of Performance on Critical Model Parameters

Figure 2 further suggests a heuristic way of thinking about semi-quantitative measures of regulatory effectiveness. Within the Markovian models appearing inside each performance state, one can compute quantities such as train unavailability. Within a simple model, train unavailability is given approximately as the product of train failure frequency times average repair time. Analogously, looking at the performance states in Figure 2, one could consider a measure of regulatory effectiveness quantified as the product of the following factors: the frequency of instances of degraded performance, the dwell time in the degraded state before the regulator successfully intervenes, and the conditional risk in the degraded state (or perhaps the change in conditional risk between normal and degraded). In order to do this, one needs to know what kinds of declining performance can occur, how frequently they occur, the resulting conditional risk, and how long it will take for a given regulatory scheme to detect them when they occur. It is easy enough to assess how long it will take to detect a given postulated performance issue, and for some such issues, it is easy enough to assess the conditional risk. Less information is currently available regarding a complete specification of kinds of performance issues, and their actual frequencies.

Figure 3 illustrates one concept of the proper relationship between the regulator and the licensee. In this particular

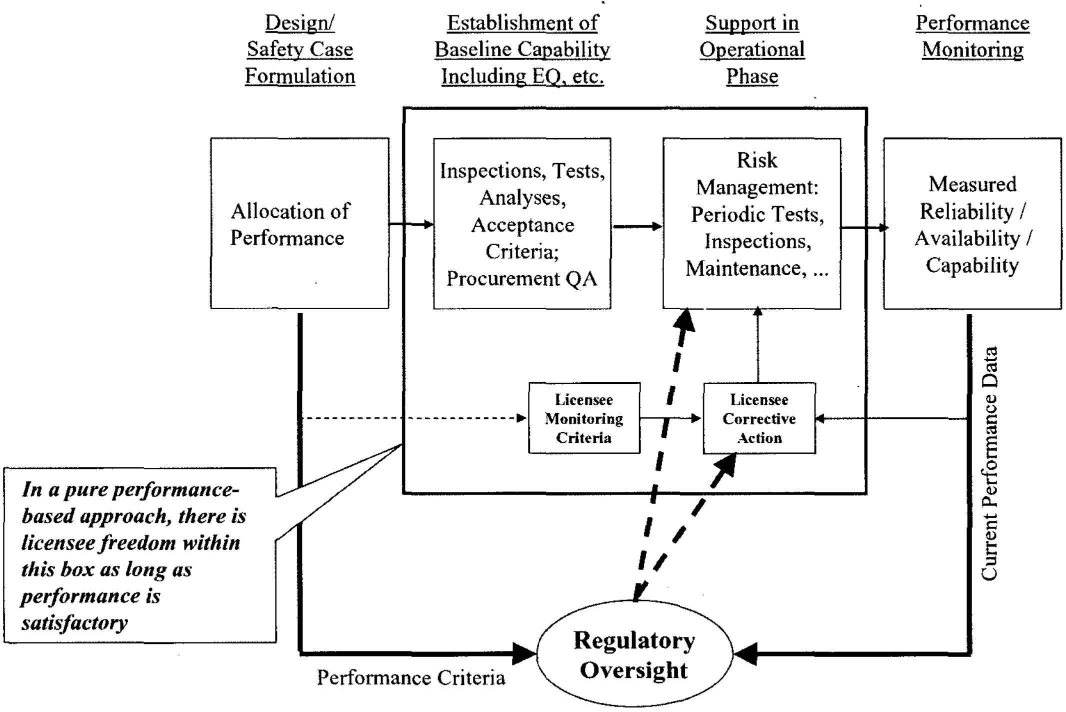

Fig. 3. Licensee Corrective Action and Regulatory Oversight in a Pure Performance-Based Approach

concept, licensees are primarily responsible for safety; it is not the regulator's role to co-manage the plant with the licensee, but rather to derive assurance that the licensee's own corrective processes are working satisfactorily, and to intervene only when there is evidence that licensee processes have not been effective. Within this concept, the licensee's own processes should first detect the descent into the "degraded" state on Figure 2, and the regulator should intervene only if the dwell time in the degraded state is long, or the degradation is especially significant. On Figure 3, the licensee's corrective action process is shown accepting current performance data, comparing current performance with licensee criteria, and responding if necessary. The internal criteria may be influenced by regulatory safety criteria, but will in general be designed to trigger licensee corrective action before regulatory action is triggered. The regulator compares current performance data with targets established in light of regulatory safety objectives, and intervenes only if necessary, e.g., licensee corrective action has not addressed the issue. This figure does not address any possible need to inspect aspects of performance that are not amenable to monitoring through reportable performance data.

The concept illustrated in Figure 3 is not unique. The idea that the regulator intervenes only when the licensee's processes have failed represents a particular policy choice, one advocated in [18] and elsewhere. A broad review of other approaches is beyond the scope of this paper, but one alternative will be mentioned briefly in Section 5.

3.2 Formal Methods for Allocation

One of the first applications of "allocation" to the development of regulatory approaches appeared in "Methods of Reliability Allocation" (NUREG/CR-4048) [22] and subsequent related publications. That work showed that it is technically feasible to allocate performance over elements of a complex system in an optimal way, or at least to identify a set of noninferior solutions, based on performance objectives related to the safety goals [23] and based on cost functions for element performance. Based on this kind of work, for a given design, and given cost functions, one could hypothetically develop a performance-based approach keyed to a set of performance targets that derive from the safety goals and are optimal from a licensee point of view. This study [22] was arguably ahead of its time; as summarized in its text, industry reviewers feared a loss of licensee flexibility if generic system-level reliability targets were promulgated based on such an approach. Also, implementation was not addressed. Today, as will be summarized later, the regulatory oversight process is analyzing the possible use of low-level plant-specific reliability targets that relate to the CDF objective, albeit through a simplified calculational process.

The NUREG/CR-4048 study treated element performance as a continuous function of element cost. Boolean optimization offers a complementary perspective. In "Top Event Prevention" (TEP) [24-27, and many other publications by the authors of 26], the question addressed is how to choose elements to be credited to satisfy performance objectives in a safety case, while optimizing cost or other metrics. Rather than fine-tuning the reliability of a particular element, TEP simply decides whether to include that element at all. TEP contemplates selection of a subset of elements (e.g., SSCs, success paths, etc.) that accomplishes the safety job as efficiently as possible. TEP can be applied at the conceptual design stage, when the question is what SSCs to include in design; or it can be applied to an existing plant, when the question is what subset of existing elements to subject to regulatory oversight in order to accomplish safety objectives in a performance-based approach.

Unlike NUREG/CR-4048 [22], TEP works at the level of detail of a modern risk model, and provides results in the form of "prevention sets." A prevention set is a complement of SSCs that satisfies the given safety objectives. Many examples of TEP have been carried out for existing plants; it is found that in many cases, there are many ways to satisfy the safety objectives, and some are clearly more efficient than others.

3.2.1 Selection of Performance Measures

Proper selection of performance measures can be a challenging task. This task is fundamentally related to central problems in decision analysis, and certain tools of decision analysis are therefore useful. To start with, the performance measures need to satisfy the attributes identified in Section 2.2. In particular, measurements need to trigger timely regulatory intervention when it is warranted.

However, in addition to this, the measures as a group should not promote undesirable outcomes, e.g., by incentivizing undesirable behavior. To take an obvious limiting example, suppose that an indicator is defined to measure component unavailability as a result of maintenance, and for the sake of the example, suppose that the purpose of the indicator is to support intervention if maintenance is excessive. Such an indicator penalizes the licensee for performing maintenance that might be necessary to achieve reliability. The situation is improved if both reliability and availability are appropriately balanced in the formulation of the indicators [28].

However, care is needed even in the definition of "unreliability," especially if an attempt is being made to work with simplified measures based on demands and failures. Some test demands may not fully test an SSC's functionality. Cases have been found in which SSC dysfunction persisted undetected through many test cycles, when failure would have occurred in the case of a real demand. Allowance needs to be made for circumstances of this kind.

3.2.2 The Objectives Hierarchy

The idea of an "objectives hierarchy" is useful in thinking about performance metrics and in understanding what makes one approach more "performance-based" than another. The subject is discussed by many authors, with some slight variation in terminology. The present discussion is based on NUREG/BR-0303 [19], which made use of [29] and [30].

An objectives hierarchy is a diagram representing the relationships and dependencies between goals, top-level fundamental objectives, lower-level fundamental objectives, and means objectives (Figure 4). Fundamental objectives are ends in themselves; means objectives are things that are desirable because they support fundamental objectives. An example of a goal is "protection of the health and safety of the public," an example of a fundamental objective is "protection of the public from excessive radiological exposures," and an example of a means objective is "reliability of safety systems."

Figure 5 shows selected elements of the objectives hierarchy appearing in SECY 99-007 [31], which developed a framework for revising the Reactor Oversight Process.

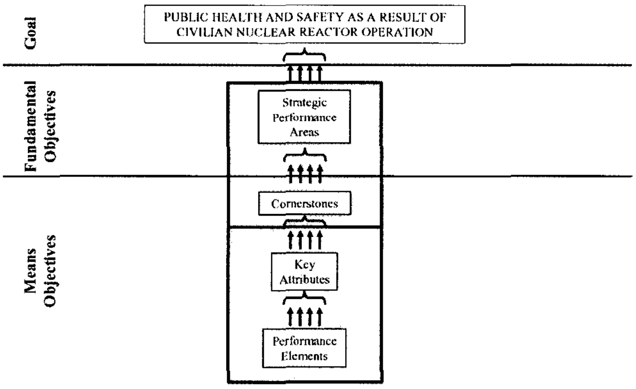

Fig. 4. Overview of Objectives Hierarchy

Seven cornerstones were identified as shown on Figure 5, and "key attributes" were identified under each cornerstone (refer to SECY 99-007 for more detail). If development of an objectives hierarchy is continued down beyond the level of "system reliability," one might identify objectives at the level of train reliability, component reliability, human actions, program implementation (testing, maintenance), and so on (Figure 6). Beginning with "system reliability," each item in this list has the property that it is important because it supports the item above, and it is in turn affected by the next item below (systems are affected by trains; components are affected by humans; humans implement established programs; and so on). Performance metrics corresponding to higher levels of the objectives hierarchy (functional reliability) are more "outcome-oriented" than performance metrics corresponding to lower levels of the objectives hierarchy (maintenance requirements). A pure performance-based approach would measure at the goal level (e.g., public safety).

Ideally, once an allocation has been carried out completely, there is a level on the objectives hierarchy at which performance targets are specified in some way for essentially all nodes. An example is specification of a complete set of system or train reliability goals in NUREG/CR-4048 [22], but an allocation can take other forms: it is logically necessary only to specify performance in such a way that satisfaction of the targets at the chosen level propagates upward to imply satisfaction of the fundamental objective. The implication of not specifying a target for a node at this level is that satisfaction of the fundamental objective is not sensitive to that node, perhaps because performance in other nodes is compensating for it. This point is discussed at length in NUREG/CR-5392 [32].

Satisfying the fourth element in the definition of "performance-based" (refer to Section 2.3) is frequently impossible if indicators and inspections are based on observations made directly at the goal level; such a framework may not trigger intervention in time to prevent unacceptable

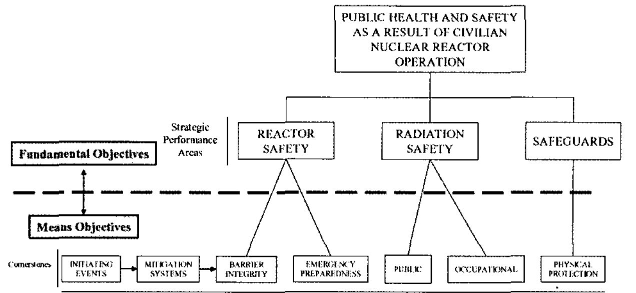

Fig. 5. Reactor Oversight Process Objectives Hierarchy

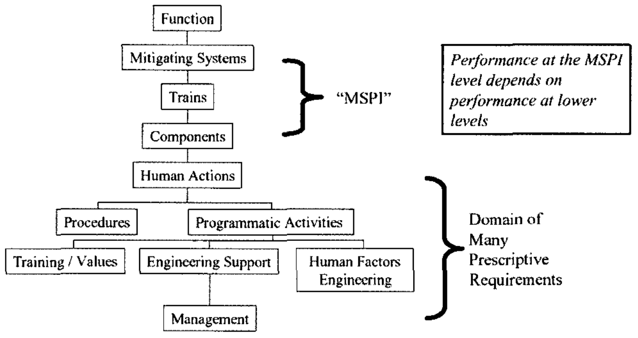

Fig. 6. Hierarchy of Performance Levels Under Mitigating Systems Cornerstone

consequences. It is practical to measure at high levels only if realistic performance issues can be identified and reversed before serious consequences result from declining performance. In general, when the consequences of accidents are serious, indicators that satisfy this property dwell several levels below fundamental objectives.

How serious must consequences be, in order to preclude direct measurement of the safety outcome as a basis for implementation? This is a policy issue, and details must be addressed case by case. For example, suppose that "serious consequences" means radiological injury or death among workers, widespread radiological contamination, or appreciable population dose to the general public. Arguably, regulatory intervention is warranted well before a declining performance trend results in such "serious" consequences.

This does not mean that performance-based regulation cannot be applied when consequences are serious; however, a systematic method for developing the regulatory approach is warranted. The essential result of the method described in NUREG/CR-5392 [32] and NUREG/BR-0303 [19] is that performance-based approaches are applicable at nodes of the objectives hierarchy that lie below the performance nodes at which failure implies "serious" consequences. In many cases, the best approach will be a blend of indicators and inspections. Note that inspections can address either prescriptive or performance-based requirements.

In setting and monitoring performance targets, it is important to recognize influences that are common to performance nodes at the same level. For example, suppose that two systems in parallel are each allocated a 10^{-2} unreliability target. Even if each appears to satisfy this target, it does not follow that their joint function satisfies a 10^{-4} target, because if they share resources or are subject to common influences, their failures are not independent. Once this consideration is recognized, coping with it is straightforward: the common element can be allowed for in the allocation, and/or can be monitored separately.

3.3 Implementation

As discussed above, in NUREG/CR-5392 [32], and in NUREG/BR-0303 [19], the implementation is the complement of measures taken to make the allocation "come true." Several kinds of implementations can usefully be discussed: prescriptive, performance-based, and process-based. All three types of requirements are found in NPP regulation. Experience suggests that it is difficult to design an optimal scheme based entirely on requirements of only one type.

As discussed in NUREG/CR-5392 [32], it may be that numerous different allocations may nominally satisfy higher-level targets, but vary significantly in their practicality or in their amenability to regulatory oversight. An important class of examples is allocations that take credit for ultra-reliable performance of systems, or system reliability at a level that is not commensurable with experience on systems having similar redundancy and diversity. Iteration between allocation and the formulation of the implementation may be necessary in order to arrive at a scheme that not only satisfies the targets on paper, but can be confirmed to do so in practice.

3.3.1 Prescriptive Implementation

In a purely prescriptive implementation, regulatory requirements prescribe what licensees should do, and licensee performance is judged by compliance with those requirements. Examples of such explicit requirements include requirements to test or inspect components on a fixed schedule, requirements governing procurement, installation, or construction of SSCs, and so on. Failure to comply with such requirements is a violation. In US NPP regulation, many of these requirements derive from consensus engineering standards.

3.3.2 Performance-Based Implementation

In a performance-based implementation, targets are established for metrics in such a way that satisfaction of the complete set of targets is deemed to correspond to accomplishment of fundamental objectives. Targets can be based on risk models, but this is not the defining characteristic of a performance-based implementation. The defining characteristics were given in Section 2.3. Inspection of Figure 6 offers a more visual notion of the relationship between prescriptive implementations and performance-based implementations: prescriptive implementation imposes requirements at the lower levels of Figure 6, while a performance-based implementation measures performance at higher levels, as the Mitigating Systems Performance Indices (MSPIs) do. The "MSPI" indicated on Figure 6 will be discussed further below.

A programmatic issue for performance-based implementations is that so far, there are no generally accepted equivalents of "compliance" and "non-compliance" with respect to performance goals. One can fail to satisfy a performance goal, but this is different from a violation of a prescriptive requirement. It is easy enough to stipulate that the regulator should intervene when performance declines to a certain level, but what form this intervention should take is difficult to specify a priori. This difficulty is circumvented to some extent by blending prescriptive and performance-based ideas. If both kinds of requirements are in force, then when declining performance is detected, it can be imputed to a compliance issue.

3.3.3 Process-Based Implementation

Several USNRC requirements essentially tell licensees to implement certain processes for decision-making within the inner loop of Figure 3 (i.e., the licensee corrective action loop between the third and fourth boxes). The Maintenance Rule [33] is such a requirement. It is frequently described as a "performance-based" requirement, but it would be more accurate to say that the Maintenance Rule is a prescriptive requirement that mandates a performance-based corrective action program within the inner loop of Figure 3. In order to qualify as performance-based, the Maintenance Rule would tell licensees what level of reliability to achieve; instead, it essentially tells them to implement a process. Failure to implement this process according to guidance is a violation, not merely failure to achieve a performance target.

4. BAYESIAN PERFORMANCE ASSESSMENT IN THE REACTOR OVERSIGHT PROCESS (ROP)

Beginning with SECY 99-007 [31], significant changes to the ROP have been undertaken. A comprehensive review of these changes is beyond the scope of the present paper, but for present purposes, certain key points are noted. Fundamental regulatory requirements are not eliminated as a result of the ROP, but regulatory inspection and enforcement is guided by the significance of adverse performance findings, assessed using risk information. The ROP makes use of both inspection information and monitored indicators. These inspections and indicators were selected using an objectives hierarchy (described above). The blend of inspections and indicators is considered to appropriately balance issues of burden with the need to trigger regulatory intervention when performance declines. Work to optimize the ROP is continuing.

The performance indicators initially formulated as part of that effort have certain drawbacks. For example, they blend unreliability and unavailability in a single train-level indicator using a method that sometimes yields a misleading picture of train-level performance. Also, the CDF significance of declining performance is not assessed plant-specifically. In response to these points, the "Risk-Based Performance Indicator" (RBPI) program[2] [34] examined alternative indicators whose formulation addressed the above shortcomings. However, the RBPIs themselves had other drawbacks; coverage of risk-significant plant systems using train-level RBPIs for unreliability and unavailability entailed using so many indicators that, using the existing decision rule for assessing performance, the statistical chance of a false-positive indication (a spurious indication of declining performance) was significantly increased as a result.

Some key features of the RBPI work were carried forward into the "Mitigating Systems Performance Index" (MSPI) program, which continues at this writing [35]. The MSPIs differ from RBPIs in that they aggregate performance information at a higher level, leading to a simpler decision process with less false-positive potential. They adopt a similar approach to assessing changes in unreliability and unavailability, and use a simple importance-based calculational approach to quantify an "index" that equates approximately to change in CDF, under certain conditions.

The problem of determining whether current reliability performance deviates from historical norms, based on sparse current data, is more difficult than estimating long-term average performance. In many problems of interest, although a significant body of historical evidence is available, current performance information is too sparse to be the sole basis for an assessment of how well the system is currently performing. For example, the "maximum likelihood estimate" (MLE) of demand unreliability (simply dividing recent failures by recent demands to obtain the current demand failure probability) yields a result that is too volatile to be used for regulatory decision-making, unless the number of demands is exceptionally high and the number of failures is also high. Therefore, it is desirable to apply current performance data within a Bayesian framework, making use of a broader body of evidence related to performance.

Correspondingly, recent work in "risk-based performance indicators" (RBPIs) and "mitigating systems performance indices" (MSPIs) for the USNRC has begun with the "constrained non-informative prior" (CNIP) distribution for failure probability (or failure rate) [36]. The assessment process is essentially to update this prior with current performance information (using Bayes' theorem), derive an estimate of change in unreliability by comparing the prior mean with the posterior mean, and use this estimate of change in unreliability in a risk-based decision rule.

The CNIP for a given failure parameter is a distribution that is formulated so as to have the industry mean of that parameter, but to maximize the entropy of the distribution, consistent with having that mean. Using current data to update the CNIP yields a posterior assessment of current performance that is influenced by past performance and therefore less volatile than the MLE, but somewhat more responsive to off-normal performance than updating a prior that is more tightly centered on the mean value. A practical advantage of the CNIP is that its mean is determined by a single parameter (the industry mean of the performance parameter). The spread in the distribution is determined by the requirement that entropy be maximized.

The CNIP was the best of several options considered in [34], but there is still potential for false-positive indications (indications of degraded performance when performance is good in reality) and false-negative indications (indications that performance is good when it is degraded in reality). Accordingly, work has been done to explore the properties of "mixture priors" for this application [37-39]. A simple mixture prior for this application can be written as

$$g_{mix}(\theta) = (1-\pi)\, g_0(\theta) + \pi\, g_1(\theta), \qquad (1)$$

---

[2] As noted earlier, indicators that are closely based on risk models can justifiably be called "risk-based" indicators, even though they are used in the context of risk-informed decision-making.

where

θ is the failure parameter (e.g., probability of failure on demand, or failure rate),
$g_0(\theta)$ is the distribution conditional on "good" performance,
$g_1(\theta)$ is the distribution conditional on degraded performance, and
π is the probability of degraded performance.

In one type of application (the "fixed-constituent" model) [38], the distributions g_0 and g_1 are considered to be fixed, and the only parameter to be updated using current data is π, the probability that performance is degraded. In another type of application (the "variable-constituent" model) [39], both constituents g_0 and g_1 are updated as well as π.

In illustrations developed for the MSPI program, g_0 and g_1 have been chosen to be conjugate distributions. The parameters of g_{mix} have been assigned based on the presumptions that (1) the prior mean should be the industry mean (as for the CNIP); (2) π, the prior probability of "degraded," should be on the order of 0.01; (3) the mean of g_1 should be 10 times the industry mean for the parameter being modeled; and (4) g_1 should be fairly diffuse.

An interesting property of the mixture prior approach is that it can directly provide an estimate of the posterior probability that performance is degraded (the posterior π). In a decision rule that focuses more closely on whether performance issues exist, and less on what the apparent risk significance is, this output is of interest in its own right. An example is shown in Figure 7. Consider the curve labeled "No Inspection Result." The quantity plotted is the posterior value of π as a function of observed failures. The prior value of π was 0.01, and the mean of the prior distribution was $5x10^{-3}$. For a small number of failures, the posterior value of π does not increase much; the assessment gives the licensee the benefit of the doubt. As the number of observed failures increases, at some point, the assessment switches over to a high posterior probability of degraded performance.

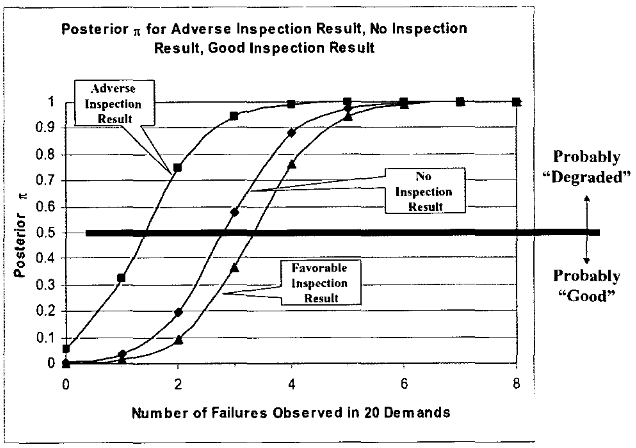

Fig. 7. Probability p that Performance is Degraded, Based on Inspections and Current Failure Data

The other curves in Figure 7 ("Favorable Inspection Result" and "Adverse Inspection Result") can be thought of in two related ways. Programmatically, they may arise if there is other information – such as inspection information – that implies either a higher π, or a lower π. Given the inspection result but before obtaining the failure data, one has a mixture distribution with a value of π updated to reflect the inspection result; the failure data are then used to update this distribution, with the result shown in the figure. Mathematically, after the inspection result is applied to the distribution, it is as if one had simply begun with a different prior value of π.

The switchover behavior shown in Figure 7 is an intuitively appealing result of using the mixture prior. Mixture priors appear to have the potential to improve the performance of the indicators by improving the responsiveness in the posterior, and can be parameterized in ways that support useful interpretations. However, work would need to be done to determine the extra parameters needed to specify the mixture priors appropriately. Most data analysis of performance parameters extracts measures of central tendency without looking for evidence of performance states, or trying to quantify their characteristics. An interesting exception is a forthcoming paper by Eide [40].

It turns out that the "fixed-state" mixture prior formalism is essentially a member of a large family of two-state decision problems, going back at least to the problem of setting alert thresholds in radar surveillance. For an interesting review of this topic, covering many applications, refer to [41]. In general terms, this family of decision problems is the following: it is necessary to decide whether an adverse condition (disease, structural flaw, performance issue) is present, based on evidence that is not completely conclusive. In general, there is a penalty for the incorrect decision: either a real adverse condition will not be addressed, or resources will be wasted in addressing an issue that is not real. In the simple form of the problem, some kind of measurement is made and compared with a threshold. The problem is to determine a decision threshold that is "optimal" in some sense. Key elements of the problem are shown conceptually on Figure 8. The horizontal axis is the range of possible observed values of the measured quantity. Two distributions are shown; one is characteristic of "no adverse condition," and the other is characteristic of "adverse condition." Conceptually, since this is a measured quantity, measurement error can contribute to the width of these distributions. If the distributions do not overlap, there is no issue: the decision is unambiguous for any given observation. If they do overlap, then there is a region in which the observation is consistent with either diagnosis.

Suppose that the vertical line on Figure 8 corresponds

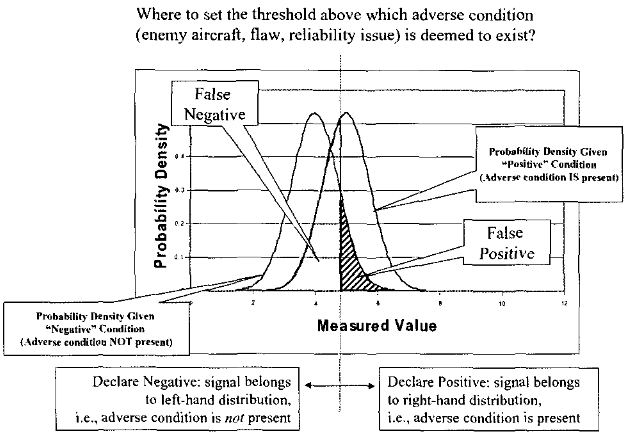

Fig. 8. Setting Decision Thresholds for Observations

to the decision threshold. This means that an observed value lying to the left of the line will imply "no adverse condition," and a value lying to the right implies "adverse condition." For the threshold shown, the diagonally hatched area shows the probability of declaring a false positive, given that there is actually no adverse condition; and the shaded area shows the probability of a false negative, given that there actually is an adverse condition. Evidently, adjusting the threshold to reduce one of the false-indication probabilities increases the other one. There is a large and still-growing literature on determination of the threshold that is optimal, given information about the distributions, the prior probability of adverse conditions, and the costs of incorrect decisions. An illustration considering the costs of misdiagnosis in ROP-type applications was given in [42], based on using failures and demands to quantify an unreliability indicator. The ROP distinguishes three classes of degraded performance; a different regulatory response is associated with each. Given a large number of demands, a large degradation in unreliability will be easy to identify; but for a moderate number of demands, a small degradation in unreliability will be difficult to determine with high confidence. The decision rule discussed above uses a threshold for change in unreliability, which (depending on the prior) translates back into a number of observed failures in a given number of demands. If there is a significant penalty associated with falsely declaring a performance issue, and only a minor penalty for temporarily accepting a small performance change, then in order to maximize utility, the decision threshold (i.e., the number of observed failures) for declaring a performance issue should be adjusted upwards. For gross changes in unreliability, the uncertainty issue is reduced, the penalties for misdiagnosis are different, and it may be appropriate to adjust the associated threshold downward.

Generally, if a given performance problem can reasonably be discussed in terms of the discrete performance states illustrated in Figure 2, and distinct conditional distributions apply to these states as in Figure 8, so that mixture priors are applicable, then insights from the literature of the two-state decision problem are also applicable. The two-state decision problem is applied widely in medical diagnosis. There is arguably a strong case for importing decision analysis concepts from medical diagnosis into the performance assessment problem: both areas involve significant uncertainties, high stakes, and competing priorities.

5. SYMPTOMATIC PERFORMANCE INDICATORS

The above discussion has focused on "risk-based" performance indicators, or more generally, performance indicators that are tied directly to nodes in the objectives hierarchy that support the fundamental objectives. However, this is not the only kind of indicator that can be considered. Formally, one can consider indicators that are determined by lower-lying performance nodes, but do not themselves directly influence risk metrics. For example, consider "Annual rate of maintenance problems (defined as maintenance rework or overdue maintenance)." Such a parameter arguably relates to basic event probabilities in a PRA model, but not in a simple way. This and several other indirect indicators are considered by the authors of [43] (and references cited therein) as what they call "Type D" indicators, meaning that they arguably correlate with important things such as safety culture, and are worth monitoring, but are not "risk-based" in the sense outlined above. (In the parlance of those authors, more model-based indicators are Types A, B, and C, in order of increasing indirectness.) Type D indicators appear to have natural analogs in medical diagnosis (symptoms that are merely cosmetic in themselves, but suggest underlying conditions that are life-threatening).

This paper does not address Type D indicators. Regulatory application of Type D indicators would be very different from implementing the double-loop characteristic of Figure 3. Figure 3 is meant to suggest an outcome-oriented regulatory regime, with criteria and monitored quantities relating directly to performance metrics of concern to the regulators. Type D indicators would make use of quantities not necessarily part of the "allocation," in the hope that such quantities would provide leading indications of declining performance. Within the double-loop paradigm, licensees might apply Type D indicators internally, but regulatory use of such indicators would be considered invasive by proponents of the double-loop approach.

6. SUMMARY

Performance-based approaches have potential advantages over prescriptive approaches:

- Performance-based approaches measure safety more directly than prescriptive approaches, giving the regulator and other stakeholders more information about the actual safety state than can be inferred from compliance with prescriptive requirements.
- Performance-based approaches provide licensees flexibility, which should allow for efficiencies and better use of resources by both licensees and regulators.

However, the application of performance-based approaches is presently limited by certain issues and development needs.

If the regulatory requirements themselves are performance-based, how can the regulator best respond when performance targets are not met?

In principle, it is straightforward to assess compliance with prescriptive requirements, and to determine penalties for violations. Performance-based requirements are another matter. Falling short of a performance goal is not necessarily the same thing as a "violation." Mechanisms of regulatory intervention need to be thought out before a truly performance-based approach to hazardous facility regulation could become feasible. An intermediate approach is to retain prescriptive requirements, but use performance-based tools in the oversight and enforcement function in order to judge the objective significance of observed violations, as now done in the ROP.

What is the best approach when severe consequences are possible in principle?

When the potential consequences of poor performance are severe, it is necessary to formulate the regulatory approach with some care. Decision analysis tools have much to offer in this undertaking. Selection of performance measures and establishment of performance criteria is appropriately done making use of an objectives hierarchy. Measurement (inspection findings, indicators) needs to target performance nodes at levels in the objectives hierarchy that are below the levels at which failure leads to severe consequences.

A study of examples suggests that in practice, the most attractive implementations will blend performance-based, prescriptive, and process-based elements. Performance-based elements will be appropriate in areas where there is margin and performance can be trended; prescriptive elements will be appropriate in areas where it is difficult to trend performance, and there is less margin; process-based elements will be appropriate when margin is present, but generic performance-based requirements are impractical to formulate or apply.

Consideration of uncertainties in performance indications can be accomplished using tools of decision analysis that have been under development for generations. Bayesian approaches are formally attractive, but as noted previously, little has been done to formulate quantitative prior expectations regarding performance issues, or to relate observable findings to conditional probabilities of performance issues. At present, use of Bayesian methods would rely heavily on judgment.

How are performance goals best determined?

The targets must be defined to be stringent enough that their collective satisfaction assures satisfaction of the high-level objectives, yet ideally, there should be some margin between expected performance and thresholds for regulatory intervention. Otherwise, regulators will intervene almost continuously, which is undesirable. Systematic formal approaches to allocation have been examined for decades and found to be technically feasible, though apart from Top Event Prevention [24-27], large-scale applications are relatively scarce.

What is needed to assess the prospective effectiveness of a proposed regulatory approach?

There are no generally accepted models that quantify the probabilities and risk consequences of programmatic lapses. In order to formulate a regulatory approach that ideally balances regulatory intervention with licensee flexibility, regulators need to understand the possible kinds and associated likelihoods (or frequencies) of programmatic lapses.

This can be understood by analogy with risk-informing itself. It is widely understood that risk modeling provides the insight needed to improve the optimization of safety resources, by considering the nature and frequency of particular challenges to safety functions. Analogous (preferably quantitative) insight is needed into performance issues in order to optimize regulatory approaches. What kinds of cross-cutting issues actually occur? How often do they occur? What do they affect? It may seem that traditional prescriptive implementation leads to the right regulatory emphasis, but there is operational evidence to suggest that this is not true. Recent operating experience suggests that under current regulatory practice, either it is possible for hazardous conditions to coexist with compliance, or noncompliance is sometimes difficult to detect, or perhaps both. There is precedent for determining inspection intervals based in part on licensee history; and from a decision-analytic point of view, it seems clear that the optimal regulatory resource allocation depends on the prior probability of performance issues of particular kinds.

Within the US at least, much reliability data analysis has been aimed at assessing industry-mean performance, and understanding long-term variability around the mean, rather than understanding how changing performance might affect the time dependence of these parameters. Recent efforts to develop performance-based elements of regulatory oversight essentially presume that licensee performance can cause reliability performance to vary; but until recently, data analysis was not typically carried out with that possibility in mind. Work on mixture priors suggests that operational experience can beneficially be viewed in a different way: it may be possible to identify some periods as "good" performance, and to identify others as "degraded performance," and to infer performance parameters for each performance state. A recent paper [40] suggests that distinct performance states (as suggested in Figure 2) may, in fact, be observable. Analytical and monitoring approaches may usefully focus on this potential.

Acknowledgment

The authors are grateful to ISL for permission to prepare this article. Although this paper draws on work carried out for various sponsors including the USNRC, this paper is not closely based on any one project. It has benefited from many conversations over the years, especially with Cory Atwood, Pat Baranowsky, Don Dube, Steve Eide, Prasad Kadambi, Steve Mays, and Scott Newberry. Opinions expressed are those of the authors, not necessarily ISL and not the USNRC.

REFERENCES

[1] U.S. Code of Federal Regulations, Title 10, Part 50.

[2] "Standard Format and Content of Safety Analysis Reports for Nuclear Power Plants, LWR Edition," Regulatory Guide 1.70, Rev. 3, USNRC (Nov. 1978).

[3] "Standard Review Plan," NUREG-0800, USNRC (July 1981).

[4] "Reactor Safety Study: An Assessment of Accident Risks in U. S. Commercial Nuclear Power Plants," WASH-1400 (NUREG 75/014), USNRC (1975).

[5] "Requirements for reduction of risk from anticipated transients without scram (ATWS) events for light-water-cooled nuclear power plants," U.S. Code of Federal Regulations, 10 CFR 50.62.

[6] "Loss of all alternating current power," U. S. Code of Federal Regulations, 10 CFR 50.63.

[7] "Additional TMI-Related Requirements," U. S. Code of Federal Regulations, 10 CFR 50.34 (f).

[8] I.S. Kim, S. Martorell, W.E. Vesely, and P.K. Samanta, "Risk Analysis of Surveillance Requirements Including Their Adverse Effects," *Reliability Engineering and System Safety*, **45**, 3, 225 (1994).

[9] P.K. Samanta, I.S. Kim, T. Mankamo, and W.E. Vesely, "Handbook of Methods for Risk-Based Analyses of Technical Specifications," NUREG/CR-6141, BNL-NUREG-52398 (Dec. 1994).

[10] I.S. Kim, P.K. Samanta, F.M. Reinhart, and M.L. Wohl, "Application of PSA to Define Technical Specifications for Advanced Nuclear Power Plants," *Proc. Int. Conf. Probabilistic Safety Assessment* (*PSA '95*), Seoul, Korea, November 26-30, 1995.

[11] I.S. Kim, "Improving Technical Specifications from a Risk Perspective," *Reliability Engineering and System Safety*, **54**, 1 (1996).

[12] "Individual Plant Examination for Severe Accident Vulne-rabilities – 10 CFR 50.54(f)," Generic Letter No. 88-20, USNRC (1988).

[13] "An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis," Regulatory Guide 1.174, Rev. 1, USNRC (2002).

[14] "An Approach for Plant-Specific, Risk-Informed Decisionmaking: Inservice Testing," Regulatory Guide 1.175, USNRC (1997).

[15] "An Approach for Plant-Specific, Risk-Informed Decisionmaking: Graded Quality Assurance," Regulatory Guide 1.176, USNRC (1997).

[16] "An Approach for Plant-Specific, Risk-Informed Decisionmaking: Technical Specifications," Regulatory Guide 1.177, USNRC (1997).

[17] "An Approach for Plant-Specific Risk-Informed Decisionmaking for Inservice Inspection of Piping," Regulatory Guide 1.178, Rev. 1, USNRC (2003).

[18] Industry White Paper, "Enhancing Nuclear Plant Safety and Reliability Through Risk-Based and Performance-Based Regulation," NEI 96-04, Nuclear Energy Institute (May 1996).

[19] N.P. Kadambi, "Guidance for Performance-Based Regulation," NUREG/BR-0303, USNRC (2002).

[20] "Risk-Informed and Performance-Based Regulation," Attachment to SECY 98-144, "White Paper on Risk-Informed and Performance-Based Regulation," USNRC (1998).

[21] E.R. Schmidt et al., "Risk Analysis and Evaluation of Regulatory Options for Nuclear Byproduct Material Systems," NUREG/CR-6642, USNRC (2000).

[22] N.Z. Cho, I.A. Papazoglou, and R.A. Bari, "A Methodology for Allocating Reliability and Risk," NUREG/CR-4048, USNRC (1986).

[23] "An Approach to Quantitative Safety Goals for Nuclear Power Plants," NUREG-0739, USNRC (1980).

[24] R.W. Youngblood and R.B. Worrell, "Top Event Prevention in Complex Systems," *Proc. 1995 Joint ASME/JSME Pressure Vessels and Piping Conf.*, PVP-Vol. 296, SERA-Vol. 3, "Risk and Safety Assessments: Where Is The Balance?" The American Society of Mechanical Engineers, New York, New York 10017 (July 1995).

[25] R.W. Youngblood, "Applying Risk Models to Formulation of Safety Cases," *Risk Analysis* **18**, 433-444, Blackwell Publishing (1998).

[26] R.B. Worrell and D.P. Blanchard, "Top Event Prevention Analysis – A Deterministic Use of PRA," *Proc. Int. Conf. Probabilistic Safety Assessment* (*PSA '95*), Seoul, Korea, November 26-30, 1995.

[27] R.A. White and D.P. Blanchard, "Development of a Risk-Informed IST Program at Palisades Using Top Event Prevention," *Proc. 10th Int. Conf. Nuclear Engineering* (*ICONE 10*), Arlington, VA, April 14-18, 2002.

[28] S. M. Wong, J. Chung, and P. Wilson, "PRA Applications and Technical Issues in NRC Maintenance Rule Baseline Inspections," *Proc. Int. Conf. Probabilistic Safety Assessment Methodology and Applications* (*PSAM 4*), New York, New York, Sept. 13-18, 1998.

[29] R.T. Clemen, *Making Hard Decisions: An Introduction to Decision Analysis*, 2nd **ed.**, Duxbury Press (1996).

[30] R. L. Keeney, *Value-Focused Thinking: A Path to Creative Decisionmaking*, Harvard University Press (1992).

[31] "Recommendations for Reactor Oversight Process Improvements," SECY 99-007, USNRC (1999).

[32] R.W. Youngblood, R.N.M. Hunt, and E.R. Schmidt, "Elements of an Approach to Performance-Based Oversight," NUREG /CR-5392, USNRC (1998).

[33] "Requirements for Monitoring the Effectiveness of Maintenance at Nuclear Power Plants," U. S. Code of Federal Regulations,

10 CFR 50.65.

[34] H.G. Hamzehee, R.W. Youngblood, S.A. Eide, et al., "Risk-Based Performance Indicators: Results of Phase I Development," NUREG-1753, USNRC (2002).

[35] D. A. Dube, C. L. Atwood, S. A. Eide, B. B. Mrowca, R. W. Youngblood, D. P. Zeek,, "Independent Verification of the Mitigating Systems Performance Index (MSPI) Results for the Pilot Plants, Final Report," NUREG-1816, USNRC (2005).

[36] C.L. Atwood, "Constrained Noninformative Priors in Risk Assessment," *Reliability Engineering and System Safety* **53**, 1, 37 (1996).

[37] C.L. Atwood and R.W. Youngblood, "Application of Mixture Priors to Assessment of Performance," *Proc. Probabilistic Safety Assessment and Management* (*PSAM 7 - ESREL'04*), June14-18, Berlin, **1**, 444 (2004).

[38] R.W. Youngblood and C.L. Atwood, "Mixture priors for Bayesian performance monitoring 1: Fixed-Constituent Model," *Reliability Engineering & System Safety* **89**, 2, 151 (2005).

[39] C.L. Atwood and R.W. Youngblood, "Mixture priors for Bayesian performance monitoring 2: Variable-Constituent Model," *Reliability Engineering & System Safety* **89**, 2, 164 (2005).

[40] S.A. Eide, "Characteristics of Nuclear Power Plant Degraded Performance," *Proc. Int. Conf. Probabilistic Safety Assessment* (*PSA '05*), San Francisco, CA, September 11-15, 2005.

[41] J.A. Swets, R.M. Dawes, and J. Monahan, "Psychological Science Can Improve Diagnostic Decisions," *Psychological Science in the Public Interest*, 1:1-26 (2000).

[42] R.W. Youngblood and C.L. Atwood, "Alternative Decision Rules for Assessing Performance," *Proc. Int. Conf. Probabilistic Safety Assessment* (*PSA '02*), Detroit, Michigan, October 6-9, 2002.

[43] M. Khatib-Rahbar, H. Erikson, and R.T. Sewell, "A New Approach to Development of a Risk-Based Safety Performance Monitoring System for Nuclear Power Plants," *Specialist Meeting on Safety Performance Indicators*, Madrid, Spain, October 17-19, 2000