INL/CON-18-52121-Revision-0

Quantitative Risk Analysis Support to Decision-Making for New Systems

Robert W Youngblood III, Homayoon Dezfuli

May 2019

The INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance

INL/CON-18-52121-Revision-0

Quantitative Risk Analysis Support to Decision-Making for New Systems

Robert W Youngblood III, Homayoon Dezfuli

May 2019

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

QUANTITATIVE RISK ANALYSIS SUPPORT TO DECISION-MAKING FOR NEW SYSTEMS

R. Youngblood,[1] H. Dezfuli[2]

[1] *Idaho National Laboratory, Idaho Falls, ID, USA; robert.youngblood@inl.gov*
[2] *H. Dezfuli, National Aeronautics and Space Administration, Washington, DC; hdezfuli@nasa.gov*

Nowadays, it is widely accepted that scenario-based probabilistic risk assessment (PRA) is needed to support key decisions about new space flight systems. However, taking a risk estimate at face value and comparing it with a management "threshold" to determine whether a system is safe enough is too simplistic. This paper discusses considerations that need to accompany PRA results in particular decision contexts. The present point is not to criticize PRA itself; others have eloquently stressed the point that, for complex, high-stakes systems, whatever PRA's imperfections, doing PRA is better than not doing it. But while all analysis results are hostage to underlying assumptions, this is especially true for PRA, and over the years, PRA results have occasionally been overinterpreted or otherwise misused. In short, there are more enlightened and less enlightened ways to understand and use PRA results. This paper discusses several issues relating to these concerns, and suggests ways of dealing with them.

I. INTRODUCTION

Compliance with classical engineering requirements on a system's physical performance can be demonstrated by running the system. However, the analog of that demonstration does not work for risk targets; it cannot in general be "proven" that an ambitious risk target is satisfied. If we want to demonstrate a very small per-launch risk of failure, we would need a very large number of launches, along with a long list of prerequisite assumptions. Purely statistical evidence from a single trial, or even a group of trials, cannot "prove" that system reliability actually meets an explicit target, except that a single success is sufficient to show that the reliability is not zero.

Instead of that sort of statistical demonstration, the decision-maker needs to rely on a safety assurance case.[1-3] The sort of case we are talking about is not absolute proof of "safety," but it is a marshaling of the relevant evidence that we have, including a clear-eyed focus on the limitations of that evidence: limitations (various types of uncertainty) that the decision-maker must understand. Nowadays, for a complex, high-stakes system, such a case will include a scenario-based model of the risk associated with the system. This paper is about the proper role of scenario-based probabilistic risk analysis (PRA) in the formulation of such a case, and certain issues associated with applying PRA results in decision-making.

Nowadays, it is widely accepted that scenario-based PRA is needed to support key decisions about new flight systems.[4] However, taking a risk estimate at face value and comparing it with a management-specified "threshold" value to determine whether a system is safe enough is too simplistic.

This being the case, why have risk targets?

- Risk targets articulate a policy tradeoff: they express the level of risk that is deemed justifiable by a given class of missions.
- Risk targets suggest:
  - how rigorous the risk analysis processes need to be (smaller risk targets imply more rigor),
  - how comprehensive and trustworthy the input evidence needs to be,
  - how hard the developers need to work in order to reduce the uncertainties.
- Risk targets say something about the level of safety at which we no longer need to sacrifice performance for safety in the design.

This paper discusses considerations that need to accompany PRA results in particular decision contexts. The present point is not to criticize PRA itself; others[5] have eloquently stressed the point that for complex, high-stakes systems, whatever PRA's imperfections, doing PRA is better than not doing it. But while all analysis results are hostage to underlying assumptions, this is especially true for PRA, and over the years, PRA results have occasionally been overinterpreted or otherwise misused. This paper discusses several issues relating to these concerns, and suggests ways of dealing with them.

II. ISSUES

II.A. Conditionality and Allocation

A naïve description of the intent of a risk model would suggest that its purpose is to quantify "the" risk. Often, however, such a model is developed in the context of a need to make a technical case that the system is adequately safe. This circumstance affects the choices that are made during the modeling effort. Any convergent modeling effort will initially try to focus on the subsystems that are going to make the most difference in the results; additionally, if the results are going to be compared with a risk target, modeling attention may focus on the systems whose performance can be modeled with the least uncertainty, because the resulting analysis will be more convincing to reviewers of the case. Finally, in a world of finite resources, when the answer seems to be good enough to make the case convincingly, additional modeling effort will be reduced.

On the other hand, sometimes, in order to attain results suggesting that the system is adequately safe, it may have been necessary to credit capabilities that are marginal in some sense. This complicates the interpretation of risk analysis results. Extremely low failure probabilities need to be viewed with skepticism, and should be presumed to have large uncertainties.

Broadly speaking, there are only certain ways to drive down the model result for a risk metric. Those ways include the following: take credit for more success paths (e.g., more redundancy) to perform a critical safety function; reduce the potentials for human error and common cause failure probability, along with the model's assessment of those potentials; incorporate in the design, and take credit for, greater operating margin; or take credit for smaller frequencies for system perturbations (e.g., initiating events) and/or lower basic event probabilities. Each of these ways may add to system cost and complexity: increased redundancy means increased capital cost and increased volume and mass (in the case of space systems); reducing common cause failure potential implies at least some redundancy, along with additional engineering effort to eliminate common influences on nominally redundant elements; smaller "independent" failure probabilities may call for increased quality assurance (QA), increased testing, and so on, and reducing assessed human error probabilities may call for all sorts of things.

In practice, we do not infallibly forecast component reliability. There is no way to absolutely guarantee a low failure probability for a given active component, even if conscientious effort is exerted to perform tests and maintenance at appropriate frequencies. The component-level equivalent of "unknown unknowns" may act on the component, or there may be a lapse in a maintenance activity. Or the actual service conditions may violate the component's engineering design basis in an unappreciated way. Correspondingly, the United States (US) Nuclear Regulatory Commission (NRC) imposes failure-tolerance requirements in safety systems of nuclear power plants. Similarly, the National Aeronautics and Space Administration (NASA) has failure-tolerance requirements for its human-rated space systems.

Some years ago, all US nuclear power plants were required[6] by the NRC to carry out "Individual Plant Examinations (IPEs)" to check for vulnerabilities to beyond-design-basis scenarios. There was no official requirement to demonstrate satisfaction of a particular target value of the risk metrics (one metric being "core damage frequency"), but most plants reported values comfortably satisfying a perceived target value for that metric (related to consistency with the Commission's Safety Goal Policy Statement[7]).

NASA has instituted requirements for establishing Agency-level safety thresholds and goals that define "long-term targeted and maximum tolerable levels of risk to the crew as guidance to space systems developers in evaluating "how safe is safe enough" for a given type of mission."[8] Safety thresholds specify the minimum tolerable/allowable level of crew safety (maximum tolerable level of risk) for the design in the context of its design reference mission, and are to be used by the Agency as criteria for program acquisition decisions.

Given all this, it is natural to ask what capabilities have been credited in the risk model in order to reach the desired risk target. For IPEs, the answer to that question was captured in a data base,[9] in terms of the success strategies invoked by the plant for each initiating event modeled, and the system-level success paths available to implement each strategy. This provided a visual indication of redundancy and diversity in system capability, which could then be roughly correlated with the risk analysis results.

For some purposes, it is useful to recast this discussion as an inverse problem. Think of the decision problem of optimizing a design: deciding what to include in the design (e.g., how much redundancy), and what levels of performance to commit to, and how to assure that those levels of performance are coming true. This is related to the problem of deciding what the PRA model *inputs* need to be, in order for the PRA's *output* risk

metrics to satisfy current objectives. Put differently: What does the PRA need to take credit for, in order to satisfy current objectives? And how much credit does it need to take? What failure probabilities can we tolerate, and what do we need to do in order to justify them? As is the case for many inverse problems, there are different methods for developing answers to such questions. One method is "Top Event Prevention Analysis," a method for finding answers to these questions that are optimal in some sense.[10-12] The decision to implement one of these answers is "allocation," discussed further below.

II.B. "Data"

There exist many sources of "data" intended for use in PRAs. But even if such data are derived from a large experience base, they do not represent constants of nature, and their applicability to a new system cannot be taken for granted. To claim a small failure probability in a specific system is to claim an engineering accomplishment. Even if a small probability of failure of a particular component seems justifiable based on operating experience, the claim implicit in using that number in an analysis is that the engineering codes and standards applied to the component being analyzed will be at least as rigorous as the codes and standards that were applied to the components whose operation gave rise to the data. Attainment and maintenance of a low level of risk imply a decision to invest to attain that level of reliability performance. This includes key design attributes, including levels of redundancy and diversity, along with other reliability assurance activities: what to include in the actual system, how to configure it, and how to operate it, in such a way that the project team actually succeeds in achieving the engineering accomplishments credited in the risk model, so that the model reflects the *actual* safety performance of the system.

These considerations operate even for a unitary decision-maker: one who is making decisions affecting only himself or herself. If the decision-maker is deciding on behalf of other stakeholders, it is even more important to carefully establish the basis for a claim of high reliability.

It is appropriate to use risk models to reason about these things, provided that the analyst thinks in terms of functional success paths, rather than individual components; but attaching credence to a low level of risk calls for appreciation of the difficulties in general of claiming low levels of risk, and the demands placed by such claims on the rigor of the methods used to argue those low levels of risk. Moreover, the difficulties of actually fulfilling those claims need to be appreciated. This is discussed below.

II.C. Performance-Based Approaches

The previous subsection discussed a thought process in which the structure of a risk model was used to allocate performance over systems, subsystems, and even components in some optimal fashion. But the allocation is only part of the story. It is then necessary to formulate the implementation: the approach to actually achieving the levels of performance that have been allocated. For example, if we are making decisions that are predicated on allocated levels of system reliability, we need for that level of system reliability to be attained in practice, so we need to think about how to ensure it. In fact, consideration of the practicality and the cost of implementation should be considered in the process of downselecting to a particular design approach.

In the past, in some domains, it was assumed that in order to achieve adequate levels of reliability, it was sufficient simply to levy prescriptive requirements on system providers to follow specific engineering practices. This may have been helpful, but it turned out not to assure highly reliable performance in all cases; and in some cases, compliance with the prescriptive requirements was found to be burdensome. Appreciation of these circumstances (ineffectiveness and burden) led to the idea of "performance-based" approaches.[13, 14] In a performance-based approach, instead of hoping that burdensome prescriptive requirements will justify the presumption of desired levels of system performance, one measures (verifies) enough about system performance to be sure of what level of performance is, in fact, being attained. It may be found that good performance is attained even if the burden of compliance with prescriptive requirements is reduced.

This is illustrated in Fig. 1, which contains a hierarchy of levels of integration at which requirements could be levied, and/or performance could be measured.

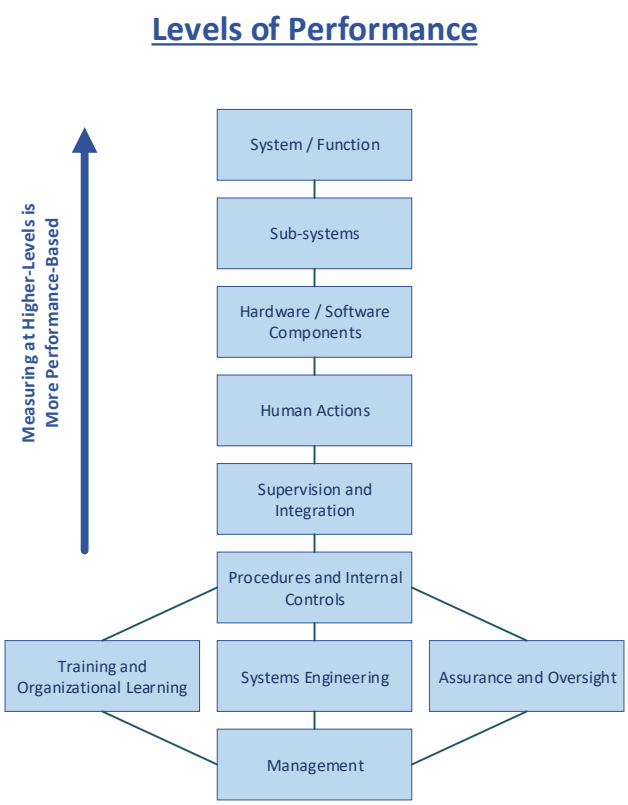

Fig. 1. Example Hierarchy of Functional Performance

Suppose that we have a firm idea of the reliability that we need at the system / function level, and we are deciding how best to allocate performance over the subsystems to achieve the desired reliability performance at the higher level. It is easy enough to default to process requirements imposed at lower levels of Fig. 1, and hope that those requirements yield subsystem performance that achieves the higher goal; but how would we know that adherence to process was actually giving us the reliability that we allocated? It would be much more satisfactory to be able to confirm the reliability of the subsystems. That way, we would have at least some information about what level of reliability is being attained at the higher level.

This is the essence of what is meant by "performance-based," and why it is a desirable approach, provided that it is practical. In general, the phrase "performance-based" is applied to implementations that are based on measuring performance at higher levels of this hierarchy, as opposed to relying on prescriptive requirements, which may not be applicable at higher levels. This is not to say that it is always practical to measure reliability at the function level: even if we can do a certain amount of functional testing, it may be impractical to actually verify functional reliability by testing a large number of times. In a large-scale system, testing may be limited to verification that the function can succeed with appropriate margin; our understanding of functional reliability will need to be derived by integrating the results of analysis, and testing at lower levels of the hierarchy. This may still be better than simply requiring providers to follow generic engineering practices and document their processes.

These considerations are discussed in several references,[13-15] including steps for formulating a maximally performance-based implementation for a given system. The bottom line is that while pure performance-based approaches are very difficult to formulate, it is straightforward to formulate implementations that combine performance-based elements with process-based elements and with more prescriptive elements, and it should generally be beneficial to do so. But the state of practice of modeling the actual benefits of performance-based approaches is still evolving.

II.D. Risk Model Incompleteness, Precursor Analysis, and Reliability Growth

Scenario-based Quantitative Risk Analysis (QRA) is "synthetic:" that is, it builds up (synthesizes) a list of risk contributors, through various processes that we will refer to here as "hazard analysis" (including everything from Hazard and Operability studies to fault-tree / event-tree analysis and beyond). This is contrasted with actuarial risk analysis, in which analysts build up statistics on occurrence rates of particular event consequences, essentially without regard to the details of the scenarios leading up to the outcomes that are captured in the statistics. An example of the latter is assessment of the risk of dying in an automobile accident; quite a bit of statistical information is available to support such an assessment. In the US, tens of thousands of people die every year in auto accidents, and this fact is often invoked in discussions of the relative safety of different transportation modalities. It is possible to parse these data in much more detail, to reflect the variation in fatality rates with geographic area, time of day, day of the week, make of car, etc. However, it is not (yet) typical to try to predict this rate by building up from scratch a detailed synthetic model of how people drive: enumerating all possible accidents, such as Driver A being distracted or drowsy during an approach to a red light or a stop sign, while another car is approaching the intersection from one side or the other when Driver A heedlessly enters the intersection, and so on. If we had no access to statistics and tried to estimate the rate of fatal accidents from such an exercise, we might later find that we had left out quite a bit, especially if we had little or no experience driving different cars in a broad range of road conditions, weather conditions, and traffic conditions, not to mention inebriation.

We have learned a great deal about the causes of automobile accidents, but a comparable level of completeness is very difficult to achieve in the development of a synthetic risk model for a novel system. We may successfully identify some previously-unknown failure modes, but we are likely to miss some others, so that our model does not reflect contributions from "unknown unknowns."

Q: Given that we cannot eliminate UU's a priori: what CAN we do?
A: Learn from operating experience as expeditiously as possible.

Published reviews of launch vehicle history (see, for example, Morse et al.[16]) and a body of experience reflected in MIL-HDBK-189[17] indicate that the first few flights of a novel launch technology are relatively risky, because of previously unknown or underappreciated failure modes. Some launch systems have done better than others, but on average, they are relatively risky early in deployment, and then improve with time (possibly with some ups and downs), as a result of changes to design and operation that result from learning from experience.

This learning process is what reliability growth modeling captures. Reliability growth modeling is not perfect, but it can be made to fit available information, and the implied narrative offers some actionable insights.

- Reliability growth modeling focuses on accomplished results viewed from the outside in, rather than trying to model launch vehicle behavior *a priori* by listing, and then modeling, failure modes.
- Because reliability growth modeling simply keeps score, all influences on reliability that operate in a particular case are implicitly taken into account, including organizational factors. Classical (fault-tree / event-tree) PRA can be made to try to account for organizational behavior, but reliability growth modeling is really about the whole complex of System + Organization + Environment + … and is moreover empirical. Classical PRA can incorporate empirical insights but is fundamentally synthetic in nature.
- Reliance on "process" to identify all failure modes has not been successful in past programs. It may be hoped that better process will do a better job of this in future developments, but it is doubtful that early-flight risk can be eliminated by better process alone. Arguably, early-flight risk can be reduced, but not eliminated, particularly for new technologies. If it were straightforward to eliminate early-flight risk, it should have been done by now.
- Reliability-growth modeling has been successful in describing program histories. After a certain number of flights, and a certain number of failures have been observed and largely eliminated, and the growth curve looks predictable, some argument can be made about current system reliability. But reliability-growth modeling is really about overall trends rather than being predictive of specific flight outcomes, and is only weakly predictive of first-flight risk. The "prediction" is this: Unless the subject system is atypical of history, there is a very real chance of one or more failures in the first several flights.

To some extent, the rate of learning from experience is under management control.[18] It cannot be increased arbitrarily, because we cannot learn from things that have not been identified in the hazard analysis and also have not happened yet. However, we can try to assure that anomalies that occur in testing or in flight are understood as completely as possible. One name for this is "precursor" analysis,[19] but this needs to be understood as analyzing anomalies in general, and not just obvious "near misses." Many failure modes will manifest initially as anomalies that signal the failure mode's existence, and create an opportunity to eliminate that failure mode, without having been severe enough to cause an accident, and without having occurred at a time or in a location where an accident would have resulted, even at the given severity. Accidents, then, are either previously-manifested failure modes that were not successfully identified or eliminated, or anomalies whose very first instantiation was severe enough to cause failure, without a previous manifestation having occurred.

In principle, precursor analysis has very significant potential to accelerate the rate of learning from experience, and thereby reducing the accident rate. It is well known[20] that precursor analysis improves the risk situation in long campaigns; arguably, precursor analysis has the potential to improve the risk situation even for short campaigns. For this purpose, "precursor analysis" needs to be formulated so as to analyze anomalies (surprises) in general,[21] and not just obvious near misses. Anomalies whose failure mechanisms were integral to the losses of Space Transportation Systems (STS) Challenger and Columbia had been occurring within the STS fleet prior to those accidents. Both the Rogers Commission Report and the Columbia Accident Investigation Board report found that processes in place at the time did not respond to the prior anomalies in a way that shed light on their true risk implications.[19] In order to help to avoid normalization of deviance, the activity needs to be predicated on a normalcy map: [22] a prior characterization of the envelope of system behaviors under normal operation, based on the current understanding, and a commitment to aggressively analyze penetrations of the normalcy envelope. A normalcy map can be a byproduct of a suitably formulated assurance case for the system, showing that outcomes are satisfactory within that envelope; what is added by this recommendation is a requirement that if the envelope is seen to be penetrated,

then the assurance case needs to be revisited, and reoccurrence of the anomaly should either be precluded, or mitigated. These measures are not guaranteed to eliminate newly identified failure modes on the first try, but on average, they should be more effective than less structured responses.

II.E. System Maturity and PRA Results

The current idea that PRA corresponds to a "mature" system result is a hope, rather than an insight. By definition, PRA cannot explicitly quantify "unknown unknowns" (UUs) (i.e., by listing them and modeling them), and therefore, its results do not apply to a system that has UU's. However, within a naïve picture, maturation is seen as a process of eliminating UUs; it is then argued that when the elimination is complete, the system is mature, and the PRA has become valid (the design has evolved to be consistent with PRA assumptions). Within this picture, the failure modes are all independent of each other, and every identified UU failure mode is eventually eliminated without the system being otherwise affected. There is no guarantee that the sources of UU risk, which are not in the PRA, will be reducible to a level of insignificance relative to the PRA result as they are uncovered and mitigated. They may be inherent in the system. Every system modification can introduce new failure modes, and it may not be practicable to eliminate any given failure mode entirely, without increasing the probability of other failure modes. This sort of reasoning suggests that PRA does not furnish an estimate of mature system performance; rather, PRA corresponds to a rough upper bound on the reliability of the "mature" system, a bound that is attained only if

1. the UU failure modes are eliminated without introducing new failure modes or exacerbating old ones,
2. the PRA faithfully models the remaining failure modes, and
3. the performance commitments implied by the PRA inputs (the basic event probabilities) are satisfied.

II.F. Treatment of Epistemic (State of knowledge) Uncertainty in Decision Analysis

The standard formalism of decision-making under uncertainty helps us decide what to do, given an unambiguous (precise) statement of uncertainty (among other things). The formalism does not eliminate uncertainty; it tells us how to reconcile our decisions with our values, given that uncertainty. It can tell us the expected value of different testing strategies, given that uncertainty. But under what conditions do we actually have a precise statement of uncertainty?

In the early 1980's, the state of practice in PRA was to formulate a diffuse prior, update it with available data, and live with the result; if data were sparse, the posterior would tend to reflect uncertainty commensurate with that of the prior, and as data accumulated, the posterior would home in on the right answer.[23] But short of having infinite data to update with, even a diffuse prior will leave its footprints in the posterior;[24] and in many cases, it is essentially impossible to formulate a "precise" statement of uncertainty.[25, 26] We are aware of being uncertain, but we cannot rigorously justify using any single, precisely specified mathematical function as a description of that uncertainty. Partly for reasons such as this, a significant and growing community of practice rejects the claim that we can justify a precise prior distribution in many cases.

The literature on "Robust Bayes" provides a sort of gateway intuition on this.[25] In Robust Bayes, instead of writing down a specific distribution as the prior, we define an ensemble of distributions that are deemed to span the set of all distributions that are reasonably consistent with our state of knowledge. Updating each member of this ensemble of distributions, we obtain an ensemble of posterior distributions; and in a favorable case, all members of the ensemble will imply the same decision. In contrast to this, the current state of practice is either to subjectively pick a distribution or assume a flat distribution between estimated upper and lower bounds on a parameter, and then average over the distribution that we picked. Unfortunately, neither current practice is really correct. For some purposes at least, we do not get to average over intervals.

Those issues relate to the treatment of epistemic uncertainty, especially the practice of proceeding as if we had a single explicit probability density distribution describing our state of knowledge regarding model parameters. Given such a distribution, for a properly-formulated decision problem, we can use the distribution to calculate "expectations:" quantities such as "expected utility of selecting a specific decision alternative" or "the expected value of reducing uncertainty regarding the model parameters." Standard decision analysis makes essential use of these ideas. But unfortunately, if we do not have explicit distributions for epistemic uncertainty, we can no longer calculate expectations; we can only bound the possibilities. This is the tip of the iceberg of "probability bounds analysis."[27]

This is not a denial of Bayes' "theorem." Bayes' theorem tells you what to do with the distributions (the priors and the likelihood models), if you have them. But by itself, Bayes' theorem doesn't guarantee that you have them. The subjectivists argue that you do have them, or have something that is close enough; but investigators working with sparse data can observe the effects of prior distributions, even ostensibly "noninformative" ones.[24]

The skeptics argue that to the degree that the uncertainty is significant, there is inherent imprecision in the formulation of the prior, and the decision-maker needs to understand the implications of this imprecision.

Some ways of addressing this issue call for working with an ensemble of probability density functions (pdfs), spanning the space of reasonable prior distributions in a context-specific way to build some prior knowledge into the ensemble of priors. Instead of specific pdfs for specific performance metrics, this leads to an ensemble of pdfs for performance metrics. So, instead of (for example) computing expected utility of a given decision alternative, we compute an ensemble of results for utility.

Unfortunately, we do not yet know how to weight the members of this ensemble: we do not know how to compute an expectation over the ensemble of results, because we do not know which members are most likely to be "right" in some sense. However, in a favorable case, the "ensemble" of posterior distributions may be so tightly clustered around a particular result that we can use the results essentially in a traditional way. Even better, all of the members of the ensemble may point to the same decision, despite variation among the ensemble members. The implications of this picture for "value of information" are, essentially, that even within the revisionist picture, it is sometimes clear that new information will add little or nothing, because the uncertainties do not cause the analysis results to straddle the effective decision criterion. But in other cases, the results will straddle the criterion. In those cases, we do not yet know how to calculate an expected value of collecting more information, but we know how to *bound* the prospective value of information collection.[28]

III. SUMMARY

Making a risk acceptance decision about a novel system is difficult: by definition, our understanding of a "novel" system cannot be based on experience with that system, and we must rely instead on modeling and analysis. Sometimes, it is tempting to reason about system adequacy based on a "worst case" scenario, but this is very often a bad idea. Historical applications of worst-case reasoning have had problems: either they push for more margin in a system than is optimal, or they miss key points that do not arise in consideration of the "worst" case. Worst-case arguments promote that which ought not to be done, and leave unconsidered that which ought to be done.

For complex, high-stakes systems, risk-informed decision-making supported by scenario-based modeling is clearly the right way to go. Scenario-based modeling provides the ability to rank risk contributors defensibly, taking account of uncertainty, and then either come up with suitable mitigations for those risks, or a suitable rationale for accepting them.

Applying the scenario-based risk model in an inverse mode has important benefits. That is, instead of specifying input parameter values and trying to apply the resulting risk result, it is useful to start with a notional threshold value of risk, and ask what levels of system and subsystem reliability need to be credited in order to satisfy that threshold value. In general, there will be more than one way to satisfy the threshold; deciding which one of those ways to implement is "allocation." Then, having decided what to take credit for, and how much credit to take, it is useful to ask what needs to be done in practice in order to achieve these levels of reliability, and confirm that they are being maintained (the "implementation"). Finally, it is useful to consider whether all these implementation measures are going to be practical, and if not, to reconsider the allocation process.

In modeling a novel system, we have no way of being sure that we have captured all of the important contributors to risk; and even for the contributors that we have captured, there may be very significant epistemic uncertainties regarding completeness, frequencies, and consequences. This limits our ability to assert compliance with probabilistic thresholds, and places a premium on maturing the system through testing and operation as quickly as possible.

ACKNOWLEDGMENTS

Work at INL was performed under DOE Idaho Operations Office Contract DE-AC07-05ID14517, with funding to INL provided through a NASA Interagency Purchase Request. We acknowledge useful conversations with numerous peers, including Curtis Smith (INL) and Chris Everett (ISL). Opinions expressed in this paper are opinions of the authors, and do not necessarily represent the views of INL or the National Aeronautics and Space Agency.

REFERENCES

1. NASA/SP-2010-580, NASA System Safety Handbook, Vol. 1, NASA, 2011.
2. NASA/SP-2014-612, NASA System Safety Handbook, Vol. 2, NASA, 2014.
3. Bishop P, Bloomfield R. A Methodology for Safety Case Development. Proceedings of the Sixth Safety-critical Systems Symposium, Feb 1998.
4. NASA/SP-2011-3421, Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners (NASA, 2011).

5. “How Useful is Quantitative Risk Assessment?”, G. E. Apostolakis, Risk Analysis 24, no. 3, pp. 515-520 (Society for Risk Analysis, 2004).
6. Generic Letter 88-20, Individual Plant Examination for Severe Accident Vulnerabilities - 10 CFR 50.54(f) (US Nuclear Regulatory Commission, November 23, 1988).
7. “Safety Goals for the Operations of Nuclear Power Plants; Policy Statement; Republication,” 51 FR 30028, August 21, 1986 (USNRC).
8. NPR 8705.2C, Human-Rating Requirements for Space Systems (NASA, 2017).
9. “IPE Data Base Structure and Insights,” J. Lehner and R. Youngblood, Water Reactor Safety Information Meeting, Bethesda, MD (United States), 25-27 Oct 1993 (USNRC, 1993).
10. "Top Event Prevention in Complex Systems," R. W. Youngblood and R. B. Worrell, Proceedings of the 1995 Joint ASME/JSME Pressure Vessels and Piping Conference, PVP-Vol. 296, SERA-Vol. 3, "Risk and Safety Assessments: Where Is The Balance?" July 1995 (The American Society of Mechanical Engineers, New York, New York 10017, 1995).
11. “Applying Risk Models To Formulation Of Safety Cases,” R. W. Youngblood, Risk Analysis 18, No. 4, p. 433 (August 1998).
12. “Top Event Prevention Analysis – A Method for Identification of Combinations of Events Important to Safety,” R. B. Worrell and D. P. Blanchard, PSA ‘02, October 2002.
13. “Elements of an Approach to Performance-Based Regulatory Oversight,” R. W. Youngblood et al., NUREG/CR-5392 (USNRC, 1999).
14. Guidance for Performance-Based Regulation,” NUREG/BR-0303 (USNRC, 2002).
15. “Issues in Formulating Performance-Based Approaches to Regulatory Oversight of Nuclear Power Plants,” R. W. Youngblood and I. S. Kim, Nuclear Engineering and Technology 37, No.3 (JUNE 2005).
16. “Modeling Launch Vehicle Reliability Growth as Defect Elimination,” Elisabeth L. Morse, Joseph R. Fragola, and Blake Putney, Proceedings of the AIAA SPACE 2010 Conference & Exposition, 30 August - 2 September 2010, Anaheim, California.
17. Department of Defense Handbook / Reliability Growth Management, MIL-HDBK-189C (Department of Defense, 2011).
18. Duffey, R. B., and Saull, J. W., Managing Risk: the human element, Wiley (2008).
19. NASA/SP-2011-3422, NASA Accident Precursor Analysis Handbook (NASA, 2011).
20. “Learning from Trending, Precursor Analysis, and System Failures,” R. W. Youngblood and R. B. Duffey, Proceedings of ABRISCO 2015 and Topical PSAM Meeting on Safety and Reliability of O&G Exploration and Production, November 23rd - 25th, 2015 - Rio de Janeiro, Brazil.
21. R. Youngblood, G. Maggio, C. Everett, and A. Hall, “Value of Analyzing Operating Events,” Proceedings of PSAM -9, Held in Hong Kong, May 2008.
22. Accident Precursor Analysis and Management / Reducing Technological Risk Through Diligence, James R. Phimister, Vicki M. Bier, Howard C. Kunreuther, Editors (National Academy Press, 2004).
23. “PRA Procedures Guide: A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants, Final Report,” Prepared under the Auspices of The American Nuclear Society and The Institute of Electrical and Electronics Engineers (U.S. Nuclear Regulatory Commission, January 1983).
24. “Effect of Epistemic Uncertainty Modelling Approach on Decision-Making: Example Using Equipment Performance Indicator,” Dana L. Kelly, Robert W. Youngblood, Proceedings of PSAM-11/ ESREL 2012, 11th International Probabilistic Safety Assessment and Management Conference and The Annual European Safety and Reliability Conference, Helsinki, Finland, 25-29 June 2012.
25. "An overview of robust Bayesian analysis, with discussion," J. O. Berger, Test 3, 5-124 (1994).
26. “Probability is perfect, but we can’t elicit it perfectly,” Anthony O’Hagan, Jeremy E. Oakley, Reliability Engineering and System Safety 85, pp. 239–248 (Elsevier, 2004).
27. “A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing,” Christopher J. Roy, William L. Oberkampf, Comput. Methods Appl. Mech. Engrg. 200, pp. 2131–2144 (Elsevier, 2011).
28. Interval Characterization of Selected Epistemic Uncertainties, Chris Everett, Homayoon Dezfuli, Bob Youngblood, and Tony Hall, 13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13), 2-7 October, 2016.