SANDIA REPORT

SAND2006-5001
Unlimited Release
Printed September 2006

Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper

Martin Pilch, Timothy G.Trucano, and Jon C. Helton

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2006-5001
Unlimited Release
Printed September 2006

Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper

Martin Pilch
Validation and Uncertainty Quantification

Timothy G. Trucano
Optimization and Uncertainty Estimation

Sandia National Laboratories
P. O. Box 5800
Albuquerque, New Mexico 87185-0370 USA

Jon C. Helton
Department of Mathematics and Statistics
Arizona State University
Tempe, Arizona 85287-1804 USA

Abstract

This report describes key ideas underlying the application of Quantification of Margins and Uncertainties (QMU) to nuclear weapons stockpile lifecycle decisions at Sandia National Laboratories. While QMU is a broad process and methodology for generating critical technical information to be used in stockpile management, this paper emphasizes one component, which is information produced by computational modeling and simulation. In particular, we discuss the key principles of developing QMU information in the form of Best Estimate Plus Uncertainty, the need to separate aleatory and epistemic uncertainty in QMU, and the risk-informed decision making that is best suited for decisive application of QMU. The paper is written at a high level, but provides a systematic bibliography of useful papers for the interested reader to deepen their understanding of these ideas.

Acknowledgements

We thank Kathleen Diegert, Art Hale, Ronald Hartwig, Scott Klenke, Laura McNamara, Scott Mitchell, Charles Nahkleh (Los Alamos National Laboratory), George Novotny, William Oberkampf, Robert Paulsen, Richard Wagner (Los Alamos National Laboratory), and Paul Yarrington for reading earlier drafts of this report, and Rhonda Reinert (Technically Write) for a critical editing of a late draft. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Contents

(Page Left Blank)

List of Figures

(Page Left Blank)

List of Acronyms

ASC	Advanced Simulation and Computing
BE+U	Best Estimate Plus Uncertainty
M&S	modeling and simulation
NRC	Nuclear Regulatory Commission
NW	Nuclear Weapons
PCMM	Predictive Capability Maturity Model
QMU	Quantification of Margins and Uncertainties
QPA	Quantitative Performance Analysis
QRA	Quantitative Risk Analysis
RIDA	Risk-Informed Decision Analysis
SNL	Sandia National Laboratories
V&V	verification and validation

(Page Left Blank)

1. Introduction

This paper presents a high-level summary of key ideas underlying one conception of Quantification of Margins and Uncertainties (QMU) for Sandia National Laboratories' (Sandia's) nuclear weapons (NW) program. Our intent is to present QMU as a methodology that, when appropriately tailored, is applicable to all the key components of the NW lifecycle. These components are generally summarized as (1) stockpile requirements, (2) design, (3) qualification, (4) production, (5) maintenance, and (6) retirement. Our perspective emphasizes modeling and simulation (M&S) and thus is not exhaustive.

QMU is generically defined by Sharp and Wood-Schultz (2003) as "a framework that captures what we do and do not know about the performance of a nuclear weapon in a way that can be used to address risk and risk mitigation." Goodwin and Juzaitis (2003) simply claim that QMU is a component in a certification methodology for the NW stockpile. The JASON study of QMU (Eardley et al., 2005) observes that the meaning and implications of QMU were still unclear to the study team at that time.

For this paper, QMU is a process used to produce information of a specific kind that is applicable to nuclear-stockpile decision making. As discussed here, stockpile decision making is about the technical management of the engineered systems in the U.S. nuclear stockpile. We are thus most concerned with requirements on the technical performance of these NW engineered systems. Performance requirements contain desirable or required performance thresholds and their associated performance margins (explained further below). QMU is thus the mathematical methodology that quantifies these thresholds and margins, as well as the associated uncertainty in their evaluation, through a "process for planning and analyzing data obtained from both tests and M&S . . ." (Klenke, 2006). QMU information is, of course, influenced in detail by the nature of the NW systems under consideration (nuclear explosive system, electronic systems, individual components, and so forth).

Given performance requirements, ***risk***[1] in stockpile stewardship can then be generally understood to be the probability or possibility (we simply say likelihood below) of failure to achieve the requirements, particularly failure to achieve the required/desired performance thresholds and margins. From our point of view, the decision making with which we are concerned must deal with this kind of risk. Developing a high-quality basis for deciding that performance will be achieved in a complex engineered system is essentially equivalent to developing a high-quality basis for understanding the likelihood that the needed performance will not be achieved. QMU provides information that helps, indeed is necessary, to quantify and understand the various performance risks in the stockpile lifecycle and that contributes to the technical basis demanded by the decision making.

Risk has a broader, more complex connotation as well—one that is not necessarily compatible with our use in QMU (see Althaus, 2005; Beck, 2004; Jasanoff, 1986; Krimsky and Golding, 1992; Slovic, 2003).[2] Rather, our use of the word "risk" is more compatible with a reliability framework for studying the likelihood of achieving or failing to achieve performance requirements expressed through thresholds and margins. The reader should keep this in mind.

QMU is, first and foremost, a decision-support methodology. Goodwin and Juzaitis (2003) discuss this from a joint Los Alamos National Laboratory–Lawrence Livermore National Laboratory perspective). The decision making that QMU supports is formally known as ***<u>risk-informed</u> decision making***. As we explain in Section 2, this risk-informed decision making accounts for other factors in addition to QMU-generated performance assessments. If decisions rest exclusively on the results of QMU assessments, then the decision process is called ***<u>risk-based</u> decision making***.

We discuss the historical basis for our conception of QMU in Section 2. We emphasize that risk-informed decision methodologies are generally accepted by regulatory bodies that are involved with complex technical policy decisions in the United States (Garrick and Christie, 2002). These bodies have generally found risk-based decision making to ultimately be lacking. Why? The short answer is that in complex decision making, risk-based decisions—essentially resting decision outcomes solely on the quantification of risk (in our case, the likelihood of failure to achieve NW performance thresholds and margins)—do not focus attention on important political, social, and economic factors that are inevitably present. Because of incomplete technical knowledge, these decisions provide only the illusion of the absence of judgment, or assertion, in the decision process. Risk-informed decision methods embrace these nontechnical factors, explicitly account for incomplete knowledge, and acknowledge the presence of judgment in the decision process.

In Sections 3 and 4, we discuss an important technical foundation for QMU, which is the formulation of key information in the form of Best Estimate Plus Uncertainty. This lies at the heart of the credibility and uncertainty-quantification methodologies that define QMU as a *technical* decision-support methodology. In Section 5, we emphasize the credibility of the M&S information that will be used in stockpile stewardship and review the particular role that rigorous verification and validation (V&V) play in establishing its credibility.

Section 6 discusses some issues that are typically raised about QMU-like activities and risk-informed decisions. We summarize our key conclusions in Section 7. A broad set of references is included to aid to the reader, as this paper does not provide a self-contained technical discussion of the issues.

2. Risk-Informed Decision Analysis for the Stockpile

QMU (***Quantitative Margins and Uncertainty***) is a decision-support methodology for complex technical decisions centering on performance thresholds and associated margins for engineered systems that are made under conditions of uncertainty. QMU supports management of the U.S. nuclear stockpile lifecycle, from driving technical requirements, through design and qualification, to production and maintenance.[3] While some have emphasized that QMU has a particularly important role in weapon performance, qualification, and stockpile assessment in the no-nuclear test era (Goodwin and Juzaitis, 2003; Sharp and Wood-Schultz, 2003), our current premise is that QMU is also needed for all phases of stockpile decision making.

QMU is not (1) a number or a set of numbers, (2) a set of functions that generates numbers, or (3) or an uncertainty-quantified analog of a set of numbers or functions. Rather, QMU is a methodology. In other words, it is a collection of methods that rest on three key elements, with the goal of supporting nuclear-stockpile decision making under uncertainty. The three key elements of our QMU methodology stress stockpile-lifecycle performance characteristics and are conveniently summarized as follows:

- <u>Element 1</u>: Identification and specification of performance threshold(s)
- <u>Element 2</u>: Identification and specification of associated performance margin(s), that is, measure(s) of exceeding performance thresholds[4]
- <u>Element 3</u>: Quantified uncertainty in threshold and margin specifications

QMU quantifies the three major elements (hence, the presence of the word “Quantitative” in QMU) and produces numbers, random variables, or some other more-general measures of uncertainty. The methodology that produces the numbers, and its formal and credible role within the larger decision context, is especially important.

An example of a performance threshold is the functioning of an electrical system. A performance threshold could be defined by the requirement to deliver an electrical signal having a minimum voltage of V_{min}. If the delivered voltage in a test of the hardware is then V_D, the margin is defined as $V_D - V_{min}$. Because there may be uncertainty in V_{min}, due to requirements uncertainties, and because V_D is uncertain, due to hardware build variations, there is likely to be a requirement that the margin should be “big enough,” such as the requirement that $V_D - V_{min} \geq M > 0$. The margin specification M itself is uncertain because of the requirements. It may be imposed basically by expert judgment, by observing the performance of a large number of built systems, or by a combination of the two. Confidence in achieving the performance threshold V_{min} in any fielded version of this electrical system should increase with the size of *M*. Other factors constrain the size of *M*, however, such as economics. One example of decision making under uncertainty related to margins and their uncertainty is how to balance increased performance confidence through increasing M versus our inability to make M large enough to completely remove performance uncertainty. Where does one draw the line? And multiplying this problem by thousands of coupled variations of similar questions gives one an idea of the overall scope and complexity of the QMU challenge that is posed by entire weapon systems. A more realistic example of attacking such a challenge is found in Helton et al. (2006) and Romero et al. (2005).

Element 3 is a crucial link to the decision process within which the QMU methodology is applied. QMU, with its careful structuring of information, its emphasis on ***Best Estimate Plus Uncertainty*** (see Section 3 below), and its rigorous attention to requirements, supports a specific paradigm of decision making under uncertainty that is highly relevant to technical policy issues. We call this paradigm ***Risk-Informed Decision Analysis*** (RIDA). RIDA acknowledges the role that human judgment plays in stockpile decisions and that technical information cannot be the sole basis for these decisions. This is because of gaps in the technical information, as well as social, economic, and political factors that inevitably influence complex national decisions. RIDA is thus compatible with an extensive body of work on the role of cognitive barriers and uncertainty in complex decision making (see Cooksey, 1996, and Hastie and Dawes, 2001, for example).

RIDA does not base its decision outcomes solely on the results of QMU. Rather, QMU provides only *part* of the input into the decision process. We might say that QMU is intended for “QMU-informed decision making” rather than for “QMU-based decision making.” We maintain our language of “risk-informed” in this paper because of its connection to nationally important technical decision-support approaches that have a long history of application. We briefly review this historical connection at the end of this section. If readers prefer to think of our concepts as more appropriately called “QMU-informed decision making,” then they should do that.

As listed below, there are good reasons to expect that the decision process that properly uses QMU would only consider QMU as one subset of the important decision variables.

- There will be uncertainty in the credibility of QMU results for complex problems, and there will be subjective information in these results.

- There will be incomplete knowledge present, for example, in the form of both known and unknown unknowns.

- Factors such as resource limitations (e.g., time constraints), as well as social, economic, and political factors that are external to the relevant scientific information, will inevitably influence the decision process.

- Complex technical decisions often rely on scientific and engineering judgment, and stockpile stewardship is no different.

That complex decision making under uncertainty necessarily entails these factors is well known (Helton, 1994). But also consider the following example. If we presume the simple formalism that in some sense QMU is simply a set of margin/uncertainty ratios (called confidence ratios by Sharp and Wood-Schultz [2003]), say, $\left\{ M_1 / U(M_1), \ldots, M_n / U(M_n) \right\}$, where M is a margin and $U(M)$ is the quantified uncertainty in that margin (see Sharp and Wood-Schultz for an illustration), then a *risk-based* decision process essentially makes a decision based only on this set of numbers. We write this kind of decision schematically as

$$\text{Decision} = D\left[M_1 / U(M_1), \ldots, M_n / U(M_n) \right].$$

This decision approach is not advocated by Sharp and Wood-Schultz. They comment on the presence of "other factors" in the decision process, such as lack of knowledge and judgment, and they prefer a decision formalization that instead looks something like

$$\text{Decision} = D\left[M_1 / U(M_1), \ldots, M_n / U(M_n); \text{other factors} \right].$$

This schematic is *still a more restricted* view of QMU than we advocate here because it still suggests that QMU is mainly a machine that produces key numbers that highly influence a decision. But, in fact, the various, important details of the uncertainty, especially how it might be reducible to the quantitative expression $U(M)$, are hidden in the construction of the ratios. "Other factors" might also be directly present in the ratios themselves, but this is far from transparent to the decision maker in this formalism. Such ratios also imply that uncertainty can be quantified in a form such as summary statistics, leaving it unclear how a methodology appropriate for separately quantifying the impact of variability and incomplete knowledge, such as the important ***probability of frequency*** approach of Kaplan and Garrick (1981),[5] can be used in the development of such ratios.

Whatever mathematical form an application of RIDA to a stockpile lifecycle decision might take, it requires that all uncertainties be identified and characterized. This includes the separate quantification of both variability (i.e., ***aleatory uncertainty***) and lack-of-knowledge uncertainty (i.e., ***epistemic uncertainty***), as well as definitions of "other factors" and quantified characterizations of their individual contributions to uncertainty. RIDA also requires attention to uncertainties in requirements and decision criteria, such as definitions of performance thresholds that are fundamental to the decision making. In addition, RIDA requires complete transparency of all the information to make the decision process understandable, traceable, and reproducible (documented).

QMU and its application in RIDA represent an evolution of the stockpile decision process favored by Sandia in the past.[6] We have depicted this evolution in Figure 1, where we suggest that RIDA lies at the top of two decision-support branches. One, labeled **C**^{3}, represents a historical theme in Sandia's stockpile decision processes. We characterize this approach as emphasizing conservative, *yet still assertion-based*, decision methods. The three elements of conservatism that are always present are (1) conservative requirements, (2) conservative scenarios, and (3) conservative assessments. This historical decision approach used by Sandia resonates strongly with test-based activities; that is, the decision process anticipates that most of the information, certainly most of the important information, is provided by test and experimental programs.

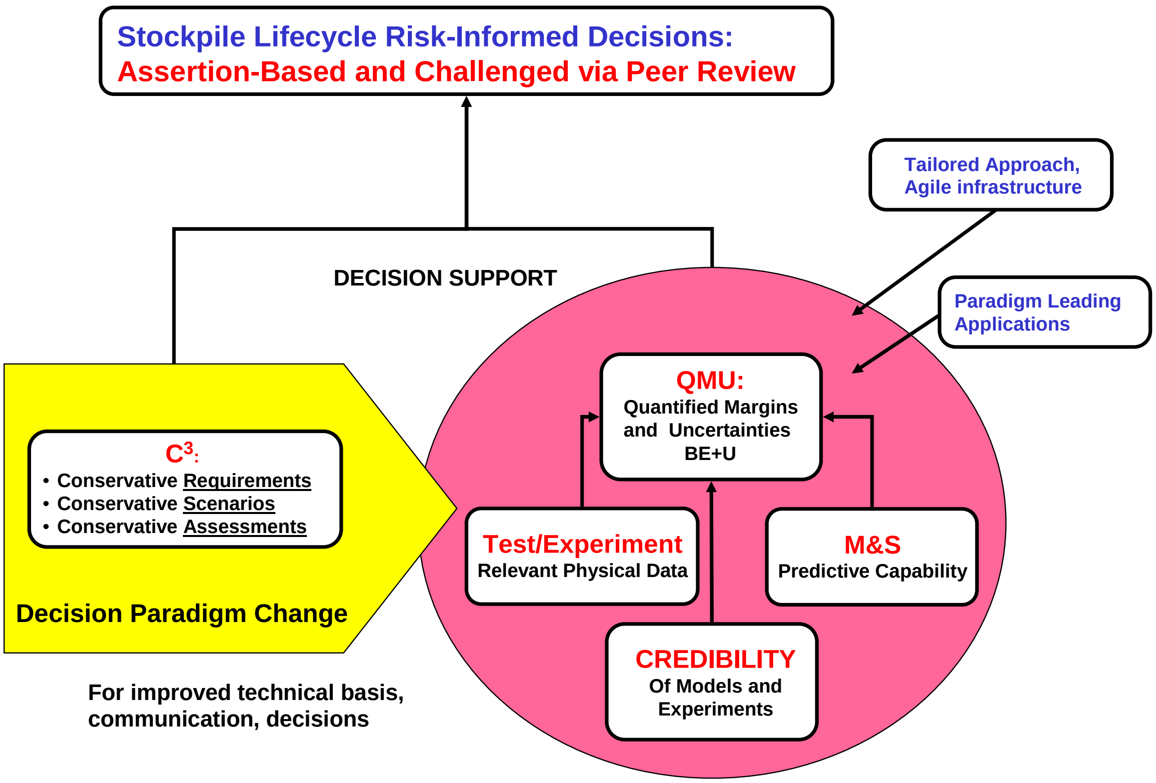

Figure 1. Elements of a decision paradigm shift that supports science-based engineering transformation through QMU and its connection to RIDA.

The product of this historical decision process typically (1) leads to the assertion of performance margins without precise quantification of the engineered margins, (2) does not directly address the scientific basis for why the relevant engineered systems behave the way they do, and (3) is not a fully effective or efficient method to understand sensitivities or uncertainties in the margin assessments or adherence to the underlying probabilistic requirements. This is primarily because the testing itself is limited, often by the social, economic, and political factors that we have previously stressed. Incomplete test information must inevitably influence decisions to include assertions. Formal and critical scrutiny of the asserted decision outcomes through aggressive peer review is therefore of major importance and has been substantially used in the past (as it would be if QMU was more extensively used).

QMU has one goal of more explicitly quantifying margins and their relationship to defined performance thresholds. This goal allows uncertainty in both quantified margins and performance thresholds, as well as the decision process itself, and clarifies the associated uncertainties, incomplete knowledge, and intangible factors that are in the decision space. Figure 1 illustrates the key principle in the presentation of this information, that is, that information is in the form of Best Estimate Plus Uncertainty (BE+U). We discuss this principle in Section 3 in greater detail.

As seen in Figure 1, QMU includes a component of M&S information that is, or will be, considerably larger than previous historical NW lifecycle decisions at Sandia. The present paper emphasizes M&S information over experimental information, although much of what we argue is true for all sources of information in RIDA. We assert that M&S information used in stockpile decisions should have a basis of credibility that is readily understandable and applicable.

The net effect of Figure 1 is to suggest evolution of Sandia's stockpile lifecycle decision processes to better account for uncertainties; better use M&S information as well as experimental information, while still relying on critical information from test and experiment; and *still* allow judgment and assertion in constructing decision outcomes. We have also suggested that other general factors, such as appropriate applications and a flexible information infrastructure, influence the overall decision process. We call this a Decision Paradigm Change in Figure 1, but our claim mainly rests on the qualitative differences in the way uncertainty is incorporated in the decision process and the (anticipated) revolutionary impact of M&S in the future. We expect that there will be cases where decision outcomes will still be test-based in the future. But even in those cases, QMU is still an important framework for integrating both testing and an enhanced scientific basis within these complex decisions. QMU guides understanding and discovery, as well as prioritization and integration of supporting efforts (such as sensitivity analysis). QMU also aids in highly risk-averse (we call this *regulatory*) decision making such as qualification and its roll-up into stockpile certification pronouncements. And QMU builds transparency for all of the above factors into the decision process. A recent example that partially illustrates the way QMU and RIDA operate on an NW issue at Sandia is presented in Helton et al. (2006) and Romero et al. (2005).

Our conception of QMU tailors an extensive historical methodological basis for quantitative performance assessment in risk-informed technical-policy decision making outside of the NW program. Sandia has led the application of such methodologies in quantitative risk assessments of nuclear power (Breeding et al., 1992; Helton and Breeding, 1993) and in development of the quantitative performance assessment for the Waste Isolation Pilot Plant (WIPP) (Helton et al., 2000a, 2000b). Sandia has recently been named the lead technical laboratory for the Yucca Mountain repository, essentially to apply the same decision methodologies there that were applied to WIPP (Helton and

Sallaberry, 2006). In our view, QMU is a customized form and application of these past RIDA methodologies to a class of decision problems that are needed for the nuclear stockpile lifecycle.

The key applications mentioned above rely upon ***Quantitative Risk Analysis*** (QRA) in the case of U.S. Nuclear Regulatory Commission nuclear reactor safety studies (the NRC NUREG-1150 study; see Breeding et al., 1992; USNRC, 1998) or ***Quantitative Performance Analysis*** (QPA) in the case of Environmental Protection Agency (EPA) waste repository assessments (associated with WIPP; see Helton et al., 2000a, 2000b) in the same sense that NW stockpile lifecycle RIDA relies upon QMU. There is no clear basis for believing that the coupling of margin and uncertainty analysis for informed decision making should philosophically differ across these decision domains. QRA, for example, produces a quantified picture of the risk of performance failures in nuclear reactor operations, while QPA does the same thing for the risk of performance failure of repositories. In neither case are decisions based solely on the results of QRA or QPA. Rather, the results are used in decisions that necessarily account for myriad other concerns. The literature associated with QRA and QPA make this clear. A detailed review of QRA/QPA is found in Rechard (1999). Garrick and Christie (2002), Niehaus and Szikszai (date unknown), and Keller and Modarres (2005) also provide useful historical context on the implementation of QRA/QPA. The main need for NW tailoring of past QRA/QPA-informed decision procedures is to account for the very different nature of the technical subject-matter disciplines that underlie the key decisions in NW stockpile lifecycle management.

3. Best Estimate Plus Uncertainty

Uncertainty enters QMU in several ways. For example, uncertainty is present in the specification of thresholds and margins. This uncertainty may arise from imprecision in the underlying requirements. Requirements uncertainty has also been common in past applications of QRA/QPA, where transformation of imprecise regulatory requirements into precise mathematical formulations is needed (see Helton, 1993; Helton and Breeding, 1993; Helton et al., 1997; and Helton, 2003). Analysis of the projected likelihood of achievement or lack of achievement of performance is based on a wide spectrum of information, including experiment and M&S, as well as myriad constraints, all of which have uncertainty. NW lifecycle decisions that result from this complex picture are necessarily examples of decision making under uncertainty.

Best Estimate Plus Uncertainty Plus Requirements (denoted simply BE+U here) is a dominant component of QMU. BE+U applies to all forms of information that are used in RIDA. For example, this specification is equally applicable to experimental information as it is to M&S information. This consistent form of information allows QMU to integrate experiment and M&S information. It also helps make transparent the sources, uncertainties, limitations, and strengths of that information. BE+U exposes the driving need for credibility of the developed information in QMU. Credibility of the information that is then used in RIDA is of special concern, and credibility is a particularly important challenge for M&S information. In our experience, neither M&S results nor test results have traditionally been developed and communicated in this form for stockpile lifecycle management.

BE (Best Estimate) is the core of understanding what we firmly know in the information. It is a function of credibility. The more we know, and the greater our belief in the credibility of that information, the more accurate and reliable BE will be. What we do not know, and the inevitable variability in a variety of components of information leading to BE, introduces uncertainty U. Both BE and U must be identified, quantified, and communicated in QMU.

We must deal with two fundamental types of uncertainty in QMU:

> ***Aleatory uncertainty*** **–** Also called "irreducible uncertainty" and "stochastic variability," this type of uncertainty is naturally identified, quantified, and communicated in terms of probability. We simply will call this type of uncertainty ***variability***. Common examples of variability are random variations in manufacturing tolerances, material composition, test conditions, and environmental factors.
>
> ***Epistemic uncertainty*** **–** Also called "reducible uncertainty," this type of uncertainty is due to incomplete knowledge. Although we may sometimes just call this ***uncertainty***, it is necessary to keep in mind that this term always refers to uncertainty due to lack of knowledge. Common examples of this type of uncertainty are model form uncertainty (that is, uncertainty about the correctness of a computational model), both known and unknown unknowns in scenarios, and poor-quality test data. It is less appreciated that an acknowledged variability that is poorly characterized stochastically due to lack of data must be treated as a lack-of-knowledge uncertainty. This kind of uncertainty may be quantified using probabilistic and statistical concepts, or other methods.

When considering QMU, the variability component of U is expected to be explicitly stochastic, with the needed statistical data underlying its quantification expected to be available. This requires the existence of a statistically significant database. Epistemic uncertainty, and its logical and mathematical distinction from variability in BE+U, is very important in stockpile stewardship, where gaps and limitations in predictive capability, incomplete experimental data, and poor statistical databases are common. Epistemic uncertainty is certainly present when there are *not enough test data to statistically quantify a presumed aleatory uncertainty*. Pilch (2005) discusses this issue, as well as the nonprobabilistic quantification of epistemic uncertainty, which is a growing area of concern. Readers should also consult Helton and Oberkampf (2004), who introduce a special issue of the journal *Reliability Engineering and System Safety* that is devoted to the quantification of epistemic uncertainty using a variety of mathematical formalisms.

Characterization, quantification, and analysis of separated aleatory and epistemic uncertainties has been the subject of a vast amount of work in the QRA/QPA technical community. A selection of useful papers that highlight the probabilistic quantification of both uncertainties in combined representations and with a view toward RIDA include Kaplan and Garrick (1981); Pate-Cornell (1986, 1996, 2002); Apostolakis (1989, 1990); Helton (1994, 1997, 1999); Helton, Johnson, and Oberkampf (2004, 2005); Ferson and Ginzburg (1996); Kadvany (1996); Winkler (1996); and the special issue of *Reliability Engineering and System Safety* devoted to this topic (Helton and Burmaster, 1996).

The traditional approach to separated aleatory and epistemic uncertainty quantification in NRC and WIPP QRAs has been a second-order probability technique (see Helton, 1993, 1994, 1997, 1999, 2003; Helton and Breeding, 1993; Helton and Burmaster, 1996; Helton et al., 1997, 2000a, 2000b; Helton, Johnson, and Oberkampf, 2004, 2005; and Helton and Sallaberry, 2006). This is also called "probability of frequency," as mentioned above.

No matter how well we achieve rigorous and credible BE+U, the remaining unknown unknowns, the gaps, and a wide spectrum of constraints force the decision making to still conform to RIDA. Specifically in the stockpile lifecycle, quantified uncertainty U is not a strict substitute for good design principles, use of safety factors, deployment of redundant systems for increased performance reliability, and application of design for computational analysis. It is important to systematically broaden uncertainty ranges beyond what is justifiable in the search for performance cliffs and unanticipated thresholds as well as other decision-threatening regions. Peer review and organizational memory are also critical.

4. QMU for System Reliability and Annual Assessment

We believe that an inviting opportunity for moving QMU forward at Sandia is to introduce this methodology into the framework for providing total system reliability estimates and using them in NW program annual assessment (Hymer and Ives, 2004). This challenging problem demands RIDA and can benefit from the increased transparency, accuracy, and completeness of the knowledge provided by systematic QMU. The NW complex is moving toward "QMUing" the nuclear-explosive packages in the enduring stockpile and is beginning to move toward "QMUing" issues in weapon system engineering. One manifestation of this is interest in a strategy to perform QMU for "critical performance parameters" reported in annual assessment, while simultaneously devising a new strategy for the Integrated Stockpile Evaluation program (SNL, 2006) applicable for future enduring systems. This state of affairs is coupled with the fact that Significant Findings Investigations may introduce additional uncertainty into the reliability numbers.

The structure that generates the system reliability "number" is *the* system-level model that embodies the sum total of all that is known about the weapon at any point in time from the decision perspective. It is at this level that prioritization and integration of supporting efforts (stated goals of QMU in support of RIDA) are best managed. We believe that all of these factors put us on a path of formally reevaluating the desirability of quantified confidence statements on reported reliability numbers.

A key question that should be answered is, How do you evolve to a new methodology without discrediting the past? For example, QMU as a decision-support tool is in potential conflict with QMU that is restricted to a tool designed for high-fidelity sensitivity analysis methodology alone. There are ways to address this concern, including (1) focusing awareness that large potential uncertainties, objectively derived, reflect current knowledge rather than assessed reliability; (2) allowing decision makers, through RIDA, to temper QMU results with other sources of input to the decision process; (3) careful management strategies, such as managing on the BE mean and managing the correlated research agenda based on recognized uncertainties; (4) not applying QMU in absolute terms (e.g., to yields) for the Stockpile Life Extension Program and, instead, applying QMU to assess the differences between old and new designs to credibly assert "no significant degradation" to a system that was previous qualified by a different methodology (e.g., nuclear testing).

While system-level reliability can serve as a good overarching goal for the application of QMU and RIDA, we recommend gaining experience with smaller, less complex issues. For example, how could additional information provided by QMU methodology intersect the current reporting of reliability during annual assessment? Reliability numbers are reported as lower-bound best estimates without meaningful confidence limits. Policy statements (most recently 1996) do not require such confidence limits. Statistical (sampling based) confidence limits can be defined, but there is no generally accepted method of measuring epistemic uncertainties and their contribution to the stated reliability numbers in these procedures. We observe that characterization of epistemic uncertainty contributions is now effectively needed for many of the issues that feed the overall system reliability model as Sandia moves into the Integrated Stockpile Evaluation program.

5. Credibility of M&S in QMU

QMU-supported RIDA has four key components of information. The first three components are analogies of the classical Kaplan-Garrick risk triple (Kaplan and Garrick, 1981) underlying the NRC/EPA applications of QRA/QPA:

I. ***Scenario identification*** **–** What can happen?

II. ***Likelihood of scenarios*** **–** How likely is it to happen?

III. ***Consequences of scenarios*** **–** What are the consequences if it does happen?

We now point out a fourth component that has always been an important factor in the use of QRAs/QPAs and that will be very important for the application of QMU in stockpile RIDA:

IV. ***Credibility*** **–** How much confidence do we have in the answers to the first three questions?

While all of these factors are important, our concern is heightened around component IV. This component is a controlling factor in the use of QMU-generated knowledge in stockpile decision processes, especially for M&S information. Our foundation for thought on this issue is formal verification and validation (V&V) of M&S. Though our focus in this section is on V&V for M&S, we emphasize that the issue of credibility of the results of QMU is equally great for experimental information as well as for information associated with social, economic, and political concerns. Uncertainty in M&S (the origin of questions about M&S credibility) has long been understood to be a critical problem in QRA/QPA (see, for example, many of the previously cited references as well as Bier, 1999, and Parry, 1996).

The credibility issue looms large for the information in RIDA that is provided by M&S. How credible is the M&S framework on which QMU must be built in the technically complex world of stockpile stewardship? In the worst possible case, if M&S is not credible, the needs of QMU cannot and will not be met. Addressing the credibility challenge for M&S is a critically important problem in QMU.

The Sandia Advanced Simulation and Computing (ASC) V&V program is tightly coupled to this issue and has been since its inception in 1999. For reviews and focused analyses of the associated challenges, the reader should consult Oberkampf and Trucano (2002); Oberkampf, Trucano, and Hirsch (2004); Pilch et al. (2004); and Trucano et al. (2002, 2003, 2006). V&V, as developed and implemented at Sandia, provides a quantitative basis for measuring the predictive credibility of M&S information entering QMU and used for RIDA-based NW lifecycle decisions.

To the degree that QMU methodology, as well as the associated RIDA decision formality, must be tailored and agile to respond to the broad qualitative differences in decision making across the stockpile lifecycle, V&V, in turn, must address the appropriate M&S credibility needs. V&V-based credibility assessment must be flexible enough to support the broad spectrum of risk aversion that we anticipate for weapon lifecycle QMU, for example, that design is less risk averse and that qualification is highly risk averse. The element of credibility was separately called out in Figure 1 as an important input into QMU. Sandia's V&V program has devoted significant attention to this issue in the past, but it remains a challenge for the future. We presented an early discussion of tailoring credibility assessment to varying levels of rigor, implying varying levels of V&V efforts, in Pilch et al. (2004). Recently, we have begun to carefully examine the development and deployment of an approach we call the ***Predictive Capability Maturity Model*** (PCMM)[7] that is even more finely aligned with the needs and applications of QMU for RIDA. (For some related ideas, see Harmon and Youngblood, 2003, 2004; and Trucano and Lott, 2003). Measuring predictive capability by some scheme also addresses an important concern of ASC, which has been programmatically challenged to answer two important questions: (1) How does one quantitatively measure and communicate progress in predictive capability? (2) How does one know when predictive capability is sufficient to achieve stated M&S requirements? Both of these are difficult problems and require a rigorous M&S predictive-capability assessment technique.

Figure 2 schematically presents the two dimensions of importance in the PCMM. The horizontal axis measures risk aversion from least to most in a notional four-bin scale. The reader should recall our brief comment in Note 2 about the complexity of risk. When we speak of risk *aversion*, we are likely dealing with some of the other dimensions of risk, such as psychological and legal, that are beyond the scope of this paper.

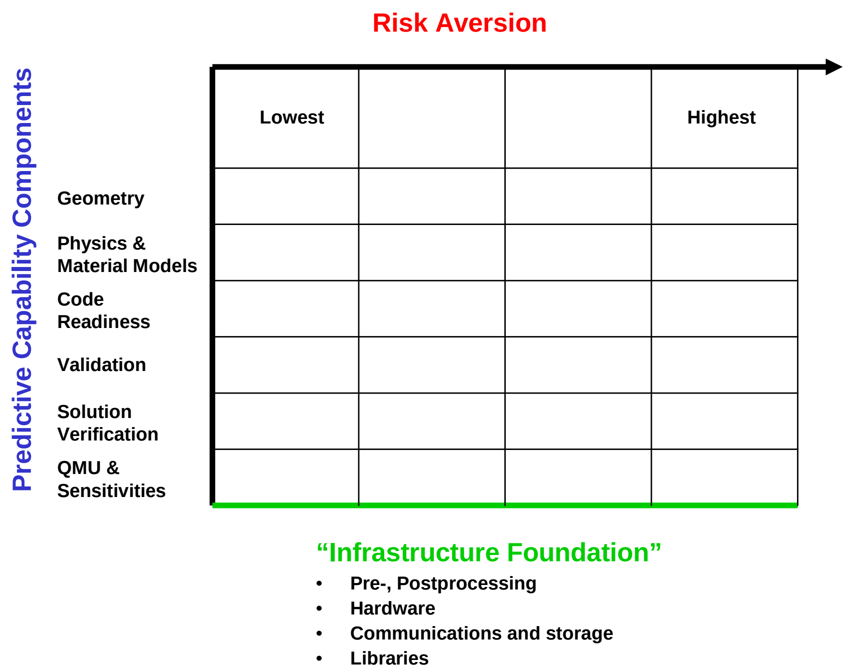

Figure 2. Key dimensions of the PCMM.

The vertical axis presents attributes of computational models that contribute to predictive capability and that should be measured. Current broad attributes of the PCMM that are intended to be quantitatively assessed include geometric fidelity, physics and material-model fidelity, code readiness (software quality engineering [SQE] and code verification), validation (separate effects and integral), solution verification (mesh adequacy for the solution of numerical partial differential equations, for example), and sensitivity analysis and other QMU-related capabilities. We also have suggested that other factors contribute to predictive capability beyond the factors addressed by the PCMM, generically labeled as the infrastructure foundation. This includes such factors as pre- and postprocessing of computational models, the hardware on which these models run, the communications and storage systems that move the associated data, and the system libraries that are required for software execution.

The broad intent of the PCMM (also compatible with discussion in Pilch et al., 2004, and Trucano and Lott, 2003) is to view increased risk aversion as demanding more effort in these separate attributes, with that effort being fundamentally measured by what is delivered. Table 1 gives an example of what we mean by delivering more as risk aversion increases. This information is intended to illustrate how the PCMM is constructed, not describe the PCMM in detail. The table suggests one approach to dealing with the sufficiency of predictive capability, presuming that there is agreement about the tasks that must be executed to achieve a given level of risk aversion in the decision process. The table also speaks to the need for balancing the technical efforts to better achieve predictive capability for a specific goal, such as QMU support in RIDA for a fixed stockpile issue. Increased rigor (that is, doing more) comes at a cost in terms of resources such as human level-of-effort, time to develop information, CPU usage on given hardware, and so on. Careful measurement in the dimensions suggested by the table should allow trade-offs to be considered, such as less geometric fidelity but more error estimation (or vice versa).

Table 1. Example of PCMM-generated tasks versus risk-aversion level.

	Lowest	Lower	Higher	Highest
Geometry	- 0D w/o significant de-featuring and stylization - 1D w/ significant de-featuring or stylization	- 1D w/o significant de-featuring and stylization - 2D w/ significant de-featuring and stylization - Appropriate 0D model justified with 1D calcs	- 2D representation of geometry w/o significant de-featuring and stylization - 3D w/ significant de-featuring and stylization - Appropriate lower-dimensional model justified with 2D calcs	- 3D rep of geometry “as built” w/o significant de-featuring or stylization - Appropriate lower-dimensional model justified with 3D model
Physics and Material Models	- Model form unknown	- Empirical model forms speculated or calibrated to represent trends - Calibration of physics-informed models	- Alternate plausible physics-informed models - Potentially w/ model form calibration	- Established physics-based model
Code Readiness	- Judgment only - Critical features and capabilities missing or lack robustness - Sustained unit/regression testing w/o significant coverage - Unsustained unit/regression testing w/ or w/o significant coverage	- Code managed and assessed against SQE requirements - Sustained unit/regression testing w/ significant coverage - Unsustained verification testing w/ or w/o significant coverage	- Code managed and assessed against SQE requirements - Sustained unit/regression testing w/ significant coverage - Sustained verification testing w/ significant coverage of separate physics	- Code managed and assessed against SQE requirements - Sustained unit/regression testing w/ significant coverage - Sustained verification testing w/ significant coverage of high-order interactions
Validation	- Judgment only - Insignificant coverage of dominant physics - Dominant physics assessed to be inadequate	- Qualitative comparisons of measurement to prediction - Substantially incomplete coverage of dominant physics	- Quantitative validation w/o assessment of variability and uncertainties in diagnostics and model - Or w/ significant extrapolation to application parameter space - W/ significant coverage of dominant physics	- Quantitative validation w/ assessment of variability and uncertainties in diagnostics and model - W/o significant extrapolation to application parameter space - W/ significant coverage of dominant physics and their interactions
Solution Verification	- Judgment only -Or numerical errors un-acceptably pollute validation or application decisions	- Explore sensitivity to discretization and algorithm parameters	- Estimate numerical errors	- Quantify rigorous numerical error bounds
QMU and Sensitivities	- Deterministic Best Estimate or nominal margins - Judgment-only assessment of uncertainty and sensitivity	- Deterministic margins - Or Informal “what if” assessment of uncertainty and sensitivity	- Initial attempts to formally quantify margins, uncertainty, and sensitivity - W/ significant judgment - Or significant judgment as to what to include	- Formal quantification of margins, uncertainty, and sensitivity - W/o significant judgment as to what to include

6. Important Challenges for QMU and RIDA in NW Lifecycle Decision Making

In this section, we present some challenges that have been raised in the literature to the use of a quantitative methodology like QMU in a decision process predicated on RIDA. We also provide some rejoinders to these challenges. General references of note for this section include Amendola (2001), Apostolakis (2004), Caruso et al. (1999), Garrick and Christie (2002), and Pate (1983).

- Subjectivity in QMU-produced knowledge (see Althaus, 2005; Bier, 1999; Cooksey, 1996; Garrick and Christie, 2002; Hastie and Dawes, 2001; Jasanoff, 1986; Niehaus and Szikszai, date unknown; Slovic, 2003).

  Excessive subjectivity threatens the usefulness and usability of the results provided by QMU. Apostolakis (2004) mentioned explicit criticism that epistemic uncertainty and the probability of extreme events cannot be realistically quantified.

  In response, we observe that subjectivity is always present in complex decisions made under conditions of uncertainty and ignorance. QMU simply emphasizes and quantifies the presence of subjectivity, while RIDA provides an appropriate framework for explicit recognition and utilization of subjectivity in the decision process. We also admit that explicit acknowledgement and incorporation of subjectivity in RIDA poses new challenges for decision makers. The presence of this inference in the results of QMU means that, in addition to ensuring the credibility of this information, we must also ensure that decision makers credibly use the information. This is part of another well-known challenge that uncertainty must not only be quantified, but quantified uncertainty must be communicated accurately and in a manner relevant to RIDA.

- Lack of firm understanding of limits of credibility of QMU (see Apostolakis, 2004; Bier, 1999).

  This factor introduces an unwillingness to use the results on the part of the decision process, whether RIDA or not.

  In response, we suggest that QMU should directly confront the credibility challenge. This is one reason that V&V is so prominent in both our thinking and our program activities centered on QMU. Sustained, rigorous V&V directly attack this challenge, at least for M&S information. Similar challenges exist for critical experimental information and other factors in the RIDA process, as well as in the formality of the decision process itself.

- Confusion of “risk-informed” with “risk-based” decision-making (Apostolakis, 2004). It is interesting that Garrick and Christie (2002) analyze the historical nature of this issue carefully, but ultimately argue in favor of moving to a more risk-based decision framework for U.S. nuclear power policy.

Another similar criticism is that despite subjectivity and credibility issues, there is a tendency to use the results of QMU (QRA/QPA historically) as the primary factor in decisions. In other words, there is mission-creep in the decision process to become risk-based, perhaps at the desire of key decision makers themselves. (Garrick and Christie would apparently disagree with this point.)

In response, we comment that an emphasis on RIDA in a properly controlled decision environment, certainly that which we expect in stockpile stewardship, will prevent this drift in decision-making fundamentals. Clearly, the criticism is patently false in the historical applications of QRA/QPA to nuclear power and environmental analyses. Specific guidance in U.S. agencies is provided to avoid this fault (see, for example, USNRC, 1998).

- Scientific inertia fosters "analysis paralysis," which impedes or defeats RIDA.

  Scientists wish to reduce epistemic uncertainty, not simply to acknowledge it, quantify it, and use it. The more scientific and technical the challenges, the greater the tendency to defer decisions while continuing to accumulate information through continued scientific research. In certain important questions (policy response to global warming, for example), this tendency is especially significant.

  In response, we contend that the following belief is also false on the basis of the historical record: that complicated policy decisions strongly coupled to technical complexity will wait an indeterminate amount of time until technical understanding is "sufficient." QMU supports decision making, that is RIDA, on time scales that are relevant to policy, and it minimizes the potential for analysis paralysis. Decisions will be made with, or without, adequate supporting information; it is best that they be made with as much supporting information as possible.

- Creeping conservatism.

  In essence, this challenge states that belief in the value of epistemic uncertainty reduction, rather than "mere" quantification and communication, by the various stakeholders in complex decisions can *force* the application of conservative assumptions and constraints in complicated decisions. Petroski (1994), for example, discusses (among other things) why bridges fail when lack-of-knowledge uncertainty is masked by conservative assumptions.

  We respond by insisting that conservative factors entering into QMU are important to avoid. It is widely believed that such factors distort the decision process (Nichols and Zeckhauser, 1988; Sielken et al., 1995; Pate-Cornell, 2002;

Diaz, 2003). Conservatism, either in the performance of QMU or in the development of constraints on its execution, also can degrade the uncertainty analysis as well as prevent meaningful sensitivity analysis (Pate, 1983, 1999). Pate-Cornell (1999) analyzes weaknesses in the Sandia QPA for WIPP that were introduced by conservative requirements originating in the EPA tasking of this analysis. This challenge is mitigated to the extent that such constraints and requirements are not part of the architecture of the QMU work. It may be difficult to avoid the imposition of such conservative requirements in strongly forced "regulatory" decision regimes.

7. Key Summary Points

The key points presented in this report are summarized as follows:

- QMU supports RIDA for stockpile lifecycle management. Stockpile lifecycle decisions use QMU-generated information, but these decisions are not solely based on that information. Expert judgment that is sensitive to other factors, including ignorance, will remain important.

- QMU emphasizes the knowledge basis of decisions, while not ignoring the myriad other social, economic, and political factors that are inevitably present in complex public-policy decisions.

- QMU highlights the characterization, organization, and communication of knowledge, whether experimental, M&S, or other. The canonical form of information provided by QMU is BE+U.

- Epistemic uncertainty—lack-of-knowledge uncertainty—is vitally important. QMU demands its acknowledgement and quantification, as well as its differentiation from aleatory uncertainty (stochastic variability).

- Rigorously assessed reliability of complete weapon systems is an important organizing principle for QMU support of stockpile lifecycle RIDA. The complexity of weapons systems drives a need for the integrated and quantified uncertainty produced by QMU. Historical approaches to RIDA for complex technical performance decisions, led by Sandia, are appropriate for tailoring and customization to the specific needs of stockpile stewardship.

- V&V is a critical component of QMU.

Notes

1. The formal specification of what we mean by risk in this paper is provided by the canonical Kaplan-Garrick risk triple (Kaplan and Garrick, 1981). This definition consists of three key components: (1) What can happen (scenarios)? (2) How likely are the scenarios? and (3) What are the consequences? We implicitly use the Kaplan-Garrick risk triple in our intuitive definition of stockpile risk: (1) What are performance thresholds and margins? (2) What is the uncertainty in these quantities and how likely is it that requirements will be achieved (or not achieved)? and (3) What are the consequences of this uncertainty? We discuss the risk triple further in Section 5, specifically in the context of M&S information.

2. Althaus (2005) reviews twelve dimensions in which risk can be analyzed: (1) linguistic and conceptual, (2) historical and narrative, (3) mathematical and logical, (4) scientific and measurable, (5) economic and decisional, (6) psychological and cognitive, (7) anthropological and cultural, (8) sociological and societal, (9) artistic and emotional, (10) philosophical and phenomenological, (11) legal and judicial, and (12) theological. We do not speak to such a complex taxonomy of risk in this paper (our attention is primarily restricted to dimensions "3", "4", and "5"), but it is worth keeping this complexity in mind to better understand the notion of risk aversion that we discuss with regard to the Predictive Capability Maturity Model in Section 5.

3. Klenke (2006) summarizes the key phases of the NW lifecycle that require QMU assessments as (1) requirements definition, (2) development and qualification, (3) production, and (4) stockpile assessment. As Klenke shows in detail, the need for performance assessment under uncertainty and attendant decisions threads through all of these phases. QMU thus provides critical information in each case.

4. We have suggested that exceeding a performance threshold defines the margin. This follows from the discussion of Sharp and Wood-Schultz (2003). We could also speak of a performance threshold as a limit that must not be exceeded. The associated margin is then the distance below the threshold of system performance thus defined. This case appears historically, for example in waste repository performance assessment (Helton et al., 2000a).

5. The ***probability of frequency*** interpretation of epistemic uncertainty stated by Kaplan and Garrick (1981) is a second-order probability interpretation of incomplete knowledge. For example, the probability of failure to achieve performance margin requirements may be defined by a poorly known frequency distribution that provides needed statistical quantities estimating the probability of failure. Uncertainty in the frequency distribution may then be quantified by introducing a probabilistic ensemble of frequency distributions, for example, by placing probability distributions on the parameters of a stated parametric frequency distribution. Estimation of the failure probability then requires estimation over the family of frequency distributions (see Kaplan and Garrick for a straightforward technical discussion). In their paper, Kaplan and Garrick also use a subjective probability interpretation of this family of frequency distributions. A more systematic treatment of this approach is found in many works by Helton (for example, see Helton, 1994).

6. QMU and RIDA methodologies are also strongly correlated with current thrusts at Sandia to "transform" the scientific basis and application of systems engineering for evolving stockpile stewardship. This is shown in the figure as SBET, or Science-Based Engineering for Transformation. A whitepaper laying out key issues in SBET is currently being developed by C. Peterson at Sandia (Peterson, 2006).

7. Our use of the word "model" in PCMM is as "a thing used as an example to follow" (*Oxford English Dictionary*). We are also purposely emulating the use of the word "model" in the well-known software Capability Maturity Model defined by the Carnegie-Mellon University Software Engineering Institute.

References

1. C. E. Althaus (2005), “A Disciplinary Perspective on the Epistemological Status of Risk,” *Risk Analysis*, Volume 25, Number 3, 567.
2. A. Amendola (2001), “Recent Paradigms for Risk Informed Decision Making,” *Safety Science*, Volume 40, 17–30.
3. G. E. Apostolakis (1989), “Uncertainty in Probabilistic Safety Assessment,” *Nuclear Engineering and Design*, Volume 115, 173–179.
4. G. E. Apostolakis (1990), “The Concept of Probability in Safety Assessment of Technological Systems,” *Science*, Volume 250, 1359–1364.
5. G. E. Apostolakis (2004), “How Useful Is Quantitative Risk Assessment,” *Risk Analysis*, Volume 24, Number 3, 515–520.
6. U. Beck (2004), *Risk Society*, Sage Publications, Thousand Oaks, California.
7. V. M. Bier (1999), “Challenges to the Acceptance of Probabilistic Risk Analysis,” *Risk Analysis*, Volume 19, Number 4, 703–710.
8. R. J. Breeding, J. C. Helton, E. D. Gorham, and F. T. Harper (1992), “Summary Description of the Methods Used in the Probabilistic Risk Assessments for NUREG-1150,” *Nuclear Engineering and Design*, Volume 135, 1–27.
9. M. A. Caruso et al. (1999), “An Approach for Using Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis,” *Reliability Engineering and System Safety*, Volume 63, 231–242.
10. R. W. Cooksey (1996), *Judgment Analysis: Theory, Methods, and Applications*, Academic Press, San Diego.
11. N. J. Diaz (2003), “Realism and Conservatism, Remarks by Chairman Diaz at the 2003 Nuclear Safety Research Conference, October 20, 2003,” NRC News, No. S-03-023, U.S. Nuclear Regulatory Commission, Washington, D.C.
12. D. Eardley et al. (2005), “Quantification of Margins and Uncertainties (QMU),” JASON Study draft report, Mitre Corporation.
13. S. Ferson and L. R. Ginzburg (1996), “Different Methods Are Needed to Propagate Ignorance and Variability,” *Reliability Engineering and System Safety*, Volume 54, 133–144.
14. B. J. Garrick and R. F. Christie (2002), “Probabilistic Risk Assessment Practices in the USA for Nuclear Power Plants,” *Safety Science*, Volume 40, 177–201.
15. B. Goodwin and R. Juzaitis (2003), “National Certification Methodology for the Nuclear Weapon Stockpile,” draft working paper (Official Use Only).
16. S. Y. Harmon and S. M. Youngblood (2003), “A Proposed Model for Simulation Validation Process Maturity,” Simulation Interoperability Workshop, Paper No. 03S-SIW-127.
17. S. Y. Harmon and S. M. Youngblood (2004), “Simulation Validation Quality and Validation Process Maturity,” Simulation Interoperability Workshop, Paper No. 04S-SIW-125.
18. R. Hastie and R. M. Dawes (2001), *Rational Choice in an Uncertain World*, Sage, New York.
19. J. C. Helton (1993), “Risk, Uncertainty in Risk, and the EPA Release Limits for Radioactive Waste Disposal,” *Nuclear Technology*, Volume 101, 18–39.

20. J. C. Helton (1994), “Treatment of Uncertainty in Performance Assessments for Complex Systems,” *Risk Analysis*, Volume 14, 483–511.
21. J. C. Helton (1997), “Uncertainty and Sensitivity Analysis in the Presence of Stochastic and Subjective Uncertainty,” *Journal of Statistical Computation and Simulation*, Volume 57, Number 1–4, 3–76.
22. J. C. Helton (1999), “Uncertainty and Sensitivity Analysis in Performance Assessment for the Waste Isolation Pilot Plant,” *Computer Physics Communications*, Volume 117, 156–180.
23. J. C. Helton (2003), “Mathematical and Numerical Approaches in Performance Assessment for Radioactive Waste Disposal: Dealing with Uncertainty,” in *Modeling Radioactivity in the Environment*, edited by E. M. Scott, 353–390, Elsevier Science, New York.
24. J. C. Helton and R. J. Breeding (1993), “Calculation of Reactor Accident Safety Goals,” *Reliability Engineering and System Safety*, Volume 39, 129–158.
25. J. C. Helton and D. E. Burmaster (1996), “Guest Editorial: Treatment of Aleatory and Epistemic Uncertainty in Performance Assessments for Complex Systems,” *Reliability Engineering and System Safety*, Volume 54, 91–94.
26. J. C. Helton et al. (1997), “Performance Assessment for the Waste Isolation Pilot Plant: From Regulation to Calculation for 40 CFR 191.13,” *Operations Research*, Volume 45, 157–177.
27. J. C. Helton et al. (2000a), “Conceptual Structure of the 1996 Performance Assessment for the Waste Isolation Pilot Plant,” *Reliability Engineering and System Safety*, Volume 69, 151–165.
28. J. C. Helton et al. (2000b), “Summary Discussion of the 1996 Performance Assessment for the Waste Isolation Pilot Plant,” *Reliability Engineering and System Safety*, Volume 69, 437–451.
29. J. C. Helton, W. L. Oberkampf, eds. (2004), “Special Issue: Alternative Representations of Epistemic Uncertainty,” *Reliability Engineering and System Safety*, Volume 85, Number 1–3, 1–376.
30. J. C. Helton, J. D. Johnson, and W. L. Oberkampf (2004), “An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions,” *Reliability Engineering and System Safety*, Volume 85, 39–71.
31. J. C. Helton, J. D. Johnson and W. L. Oberkampf (2006), “Probability of Loss of Assured Safety in Temperature Dependent Systems with Multiple Weak and Strong Links,” *Reliability Engineering and System Safety*, Volume 91, 320–348.
32. J. C. Helton and C. J. Sallaberry (2006), “Illustration of Sampling-Based Approaches to the Calculation of Expected Dose in Performance Assessments for the Proposed High Level Radioactive Waste Repository at Yucca Mountain, Nevada,” to be published.
33. R. L. Hymer and E. E. Ives (2004), “Review of the Stockpile Surveillance Program of the National Nuclear Security Administration (NNSA),” report to Assistant Deputy Administrator, NNSA, for Military Applications and Stockpile Operations (Official Use Only).
34. S. Jasanoff (1986), *Risk Management and Political Culture*, Sage Publications, Thousand Oaks, California.

35. J. Kadvany (1996), "Taming Chance: Risk and the Quantification of Uncertainty," *Policy Science*, Volume 29, 1–27.
36. S. Kaplan and B. J. Garrick (1981), "On the Quantitative Definition of Risk," *Risk Analysis*, Volume 1, Number 1, 11–27.
37. W. Keller and M. Modarres (2005), "A Historical Overview of Probabilistic Risk Assessment Development and Its Use in the Nuclear Power Industry: A Tribute to the Late Professor Normal Carl Rasmussen," *Reliability Engineering and System Safety*, Volume 89, 271–285.
38. S. Klenke (2006), "Quantification of Margins and Uncertainties (QMU) in the Nuclear Weapons Lifecycle," Sandia National Laboratories (draft) design guide.
39. S. Krimsky and D. Golding, eds. (1992), *Social Theories of Risk*, Praeger, Westport, Connecticut.
40. A. L. Nichols and R. J. Zeckhauser (1988), "The Perils of Prudence: How Conservative Risk Assessments Distort Regulation," *Regulatory Toxicology and Pharmacology*, Volume 8, 61–75.
41. F. Niehaus and T. Szikszai (date unknown), "Risk Informed Decision Making," unpublished manuscript.
42. W. L. Oberkampf and T. G. Trucano (2002), "Verification and Validation in Computational Fluid Dynamics," *Progress in Aerospace Sciences*, Volume 38, 209–272.
43. W. L. Oberkampf, T. G. Trucano, and C. Hirsch (2004), "Verification, Validation, and Predictive Capability in Computational Engineering and Physics," *Applied Mechanics Reviews*, Volume 57, Number 5, 345–384.
44. G. W. Parry (1996), "The Characterization of Uncertainty in Probabilistic Risk Assessments of Complex Systems," *Reliability Engineering and System Safety*, Volume 54, 119–126.
45. M. E. Pate (1983), "Acceptable Decision Processes and Acceptable Risks in Public Sector Regulations," *IEEE Transactions on Systems, Man, and Cybernetics*, Volume SMC-3, 113–124.
46. M. E. Pate-Cornell (1986), "Probability and Uncertainty in Nuclear Safety Decisions," *Nuclear Engineering and Design*, Volume 93, 319–327.
47. M. E. Pate-Cornell (1996), "Uncertainties in Risk Analysis: Six Levels of Treatment," *Reliability Engineering and System Safety*, Volume 54, 95–111.
48. M. E. Pate-Cornell (1999), "Conditional Uncertainty Analysis and Implications for Decision-Making: The Case of WIPP," *Risk Analysis*, Volume 19, Number 5, 995–1002.
49. M. E. Pate-Cornell (2002), "Risk and Uncertainty Analysis in Government Safety Decisions," *Risk Analysis*, Volume 22, Number 3, 633–646.
50. H. Petroski (1994), *Design Paradigms*, Cambridge University Press, Cambridge, England.
51. M. Pilch et al. (2004), "Concepts for Stockpile Computing," Sandia National Laboratories, SAND2004-2479. (Official Use Only)
52. M. Pilch (2005), "The Method of Belief Scales as a Means for Dealing with Uncertainty in Tough Regulatory Decisions," Sandia National Laboratories, SAND2005-4777. (Available at http://www.sandia.gov)

53. R. P. Rechard (1999), "Historical Relationship Between Performance Assessment for Radioactive Waste Disposal and Other Types of Risk Assessment," *Risk Analysis*, Volume 19, Number 5, 763–807.
54. V. J. Romero et al. (2005), "Advances in an Approach to QMU (Quantitative Margins and Uncertainty) Applied to Weapon Safety in Abnormal Thermal Environments," Sandia National Laboratories, SAND2005-1322 (Official Use Only).
55. Sandia National Laboratories (2006), "Integrated Stockpile Evaluation: Transforming the Nuclear Weapons Stockpile Evaluation Program" (Official Use Only).
56. D. H. Sharp and M. M. Wood-Schultz (2003), "QMU and Nuclear Weapons Certification: What's Under the Hood," *Los Alamos Science*, Number 28, 47–53.
57. R. L. Sielken, Jr., R. S. Bretzlaff, and D. E. Stevenson (1995), "Challenges to Default Assumptions Stimulate Comprehensive Realism as a New Tier in Quantitative Cancer Risk Assessment," *Regulatory Toxicology and Pharmacology*, Volume 21, 270–280.
58. P. Slovic (2003), "Going Beyond the Red Book: The Sociopolitics of Risk," *Human and Ecological Risk Assessment*, Volume 9, 1–10.
59. T. G. Trucano, M. Pilch, and W. L. Oberkampf (2002), "General Concepts for Experimental Validation of ASCI Code Applications," Sandia National Laboratories, SAND2002-0341. (Available at http://www.sandia.gov)
60. T. G. Trucano, M. Pilch, and W. L. Oberkampf (2003), "On the Role of Code Comparisons in Verification and Validation," Sandia National Laboratories, SAND2003-2752. (Available at http://www.sandia.gov)
61. T. G. Trucano and S. E. Lott (2003), "Formality Levels in Computational Engineering Requirements Management," presentation at the 5th Biennial Tri-Laboratory Engineering Conference, October 21–23, 2003, Santa Fe, New Mexico. (Available at http://www.sandia.gov)
62. T. G. Trucano et al. (2006), "Calibration, Validation, and Sensitivity Analysis: What's What," *Reliability Engineering and System Safety*, Volume 91, 1331–1357..
63. U.S. Nuclear Regulatory Commission (USNRC) (1998), "An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant Specific Changes to the Licensing Basis, Regulatory Guide, 1.174," Washington D.C.
64. R. L. Winkler (1996), "Uncertainty in Probabilistic Risk Assessment," *Reliability Engineering and System Safety*, Volume 54, 127–132.

Sandia Internal Distribution

1	MS 0102	0250	R. K. Wilson
1	MS 1415	1120	C. J. Barbour
1	MS 1179	1341	L. Lorence
1	MS 1146	1384	P. J. Griffin
1	MS 0321	1400	W. J. Camp
1	MS 1110	1400	D. E. Womble
1	MS 0310	1410	M. D. Rintoul
1	MS 0370	1411	S. A. Mitchell
1	MS 0370	1411	B. Adams
1	MS 0370	1411	R. A. Bartlett
1	MS 1110	1411	D. Dunlavy
1	MS 0370	1411	M. S. Eldred
1	MS 0370	1411	D. M. Gay
1	MS 1110	1411	J. Hill
1	MS 1111	1411	P. Knupp
1	MS 0370	1411	L. P. Swiler
1	MS 0370	1411	S. Thomas
10	MS 0370	1411	T. G. Trucano
1	MS 0370	1411	B. G. van Bloemen Waanders
1	MS 1110	1412	S. J. Plimpton
1	MS 1110	1414	S. S. Collis
1	MS 1110	1414	M. Heroux
1	MS 1110	1414	R. B. Lehoucq
1	MS 1110	1415	S. K. Rountree
1	MS 1110	1415	W. E. Hart
1	MS 1111	1415	B. A. Hendrickson
1	MS 1110	1415	C. A. Phillips
1	MS 1111	1416	A. G. Salinger
1	MS 0316	1420	S. S. Dosanjh
1	MS 1109	1420	J. Tompkins
1	MS 0376	1421	T. D. Blacker
1	MS 0817	1422	J. A. Ang
1	MS 0817	1422	R. Benner
1	MS 1110	1423	N. D. Pundit
1	MS 0321	1430	J. E. Nelson
1	MS 0378	1431	R. M. Summers
1	MS 0370	1431	M. E. Kipp
1	MS 0378	1431	A. C. Robinson
1	MS 0378	1431	S. A. Silling
1	MS 0370	1431	G. Weirs
1	MS 0378	1433	J. Strickland
1	MS 0370	1433	G. Backus
1	MS 0310	1433	M. Boslough
1	MS 0370	1433	L. A. McNamara
1	MS 0316	1437	S. A. Hutchinson
1	MS 0316	1437	J. Castro
1	MS 0316	1437	J. N. Shadid
1	MS 0384	1500	A.C. Ratzel
1	MS 0826	1500	D. K. Gartling
1	MS 0824	1500	T.Y. Chu

1	MS 0836	1500	M.R. Baer
1	MS 0825	1510	W. Hermina
1	MS 0834	1512	R. D. M. Tachau
1	MS 0836	1514	J. S. Lash
1	MS 0825	1515	B. Hassan
1	MS 0836	1516	E. S. Hertel
1	MS 0836	1516	D. Dobranich
1	MS 0836	1516	R. E. Hogan
1	MS 0836	1517	R. O. Griffith
1	MS 0847	1520	P. J. Wilson
1	MS 0555	1522	R. A. May
1	MS 0847	1523	T. J. Baca
1	MS 0372	1524	J. Pott
1	MS 0372	1525	J. Jung
1	MS 0847	1526	J. M. Redmond
1	MS 0847	1526	R. V. Field
1	MS 0821	1530	A.L. Thornton
1	MS 1135	1532	S. R. Tieszen
10	MS 0828	1533	M. Pilch
1	MS 0828	1533	A. R. Black
1	MS 0828	1533	K. J. Dowding
1	MS 0828	1533	A. A. Giunta
20	MS 0779	1533	J. C. Helton
5	MS 0828	1533	W. L. Oberkampf
1	MS 0557	1533	T. L. Paez
1	MS 0828	1533	J. R. Red-Horse
1	MS 0828	1533	V. J. Romero
1	MS 0828	1533	A. Urbina
1	MS 0828	1533	W. R. Witkowski
1	MS 1135	1534	S. Heffelfinger
1	MS 0384	1540	H. S. Morgan
1	MS 0380	1542	K. F. Alvin
1	MS 0382	1543	J. R. Stewart
1	MS 1152	1652	M. L. Kiefer
1	MS 1186	1674	T. A. Mehlhorn
1	MS 1186	1674	C. J. Garasi
1	MS 0139	1900	A. Hale
1	MS 0139	1902	P. Yarrington
1	MS 0139	1904	R. K. Thomas
1	MS 0457	2001	G. C. Novotny, Jr.
1	MS 0429	2100	B. C. Walker
1	MS 0429	2100	R. C. Hartwig
1	MS 0453	2110	L. S. Walker
1	MS 0447	2111	J. D. Mangum
1	MS 0483	2112	A. L. Hillhouse
1	MS 0427	2118	R. A. Paulsen
1	MS 0427	2118	S. E. Klenke
1	MS 0427	2118	B. H. Wohl
1	MS 0453	2120	M. R. Sjulin
1	MS 0482	2123	E. R. Hoover
1	MS 0487	2124	P. A. Sena
1	MS 0453	2130	M. A. Rosenthal
1	MS 0481	2132	S. G. Barnhart
1	MS 0481	2137	J. F. Nagel

1	MS 0479	2138	J. O. Harrison
1	MS 0529	2345	G. K. Froehlich
1	MS 0512	2500	T. E. Blejwas
1	MS 1064	2614	S. E. Lott
1	MS 0437	2820	K. D. Meeks
1	MS 0437	2830	J. M. McGlaun
1	MS 0807	2990	P. A. Klein
1	MS 0509	5345	M. W. Callahan
1	MS 0831	5500	M. O. Vahle
1	MS 0776	6853	R. P. Rechard
1	MS 0736	6870	D. A. Powers
1	MS 0839	7000	G. Yonas
1	MS 9202	8205	R. M. Zurn
1	MS 9014	8242	A. R. Ortega
1	MS 9042	8774	J. J. Dike
1	MS 9153	8200	C. L. Knapp
1	MS 9153	8800	D. R. Henson
1	MS 9159	8962	H. R. Ammerlahn
1	MS 9159	8962	P. D. Hough
1	MS 9159	8962	M. L. Martinez-Canales
1	MS 0428	12330	T. R. Jones
1	MS 0434	12334	B. M. Mickelsen
1	MS 0830	12335	K. V. Diegert
1	MS 0829	12337	J. M. Sjulin
1	MS 0829	12337	B. M. Rutherford
1	MS 0428	12340	V. J. Johnson
1	MS 0638	12341	D. E. Peercy
1	MS 0110	12900	M. J. Cieslak
1	MS 0437	12930	J. M. McGlaun
2	MS 9018	08944	Central Technical Files
2	MS 0899	04536	Technical Library