

QUANTIFICATIONS OF MARGINS AND UNCERTAINTIES (QMU)

STUDY LEADER

D. Eardley

CONTRIBUTORS:

H. Abarbanel	J. Katz
J. Cornwall	S. Koonin
P. Dimotakis	D. Long
S. Drell	D. Meiron
F. Dyson	R. Schwitters
R. Garwin	J. Sullivan
R. Grober	C. Stubbs
D. Hammer	P. Weinberger
R. Jeanloz	

J. Kammerdiener (Consultant)

JSR-04-330

March 23, 2005

Distribution authorized to the Department of Defense and the National Nuclear Security Administration and its contractor laboratories (NA-11). Other requests for this document shall be referred to the National Nuclear Security Administration.

JASON
The MITRE Corporation
7515 Colshire Drive
McLean, Virginia 22102
703-883-6997

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information estimated to average 1 hour per response, including the time for review instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

March 2005

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE

Quantifications of Margins and Uncertainties (QMU)

5. FUNDING NUMBERS

13059022-IN

6. AUTHOR(S)

D. Eardley, et al.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The MITRE Corporation
 JASON Program Office
 7515 Colshire Drive
 McLean, Virginia 22102

8. PERFORMING ORGANIZATION REPORT NUMBER

JSR-04-330

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Department of Energy
 National Nuclear Security Administration
 Washington, DC 20585

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

JSR-04-330

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution authorized to the Department of Defense and the National Nuclear Security Administration and its contractor laboratories (NA-11). Other requests for this document shall be referred to the National Nuclear Security Administration.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Quantification of Margins and Uncertainties is a formalism for dealing with the reliability of complex technical systems, and the confidence which can be placed in estimates of that reliability. We are specifically concerned with its application to the performance and safety of the nuclear stockpile, because the test moratorium precludes direct experimental verification. We define performance "gates", margins and uncertainties and discuss how QMU differs from conventional error propagation. Finally, we review the history of QMU and its meaning, and explore how it may evolve to meet future needs.

14. SUBJECT TERMS

15. NUMBER OF PAGES

35

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT

UNCLASSIFIED

20. LIMITATION OF ABSTRACT

SAR

Contents

1 EXECUTIVE SUMMARY	1
1.1 Observations, Conclusions and Recommendations	1
1.2 What Must Be Done in the Next Year or Sooner	3
2 INTRODUCTION	5
3 QMU PRIMER	7
3.1 Gates, Cliffs, Error Propagation and QMU	8
3.2 Uncertainties	12
3.2.1 Indeterminacy (Chaos)	13
3.2.2 Human uncertainties	14
4 AN HISTORICAL VIEW OF QMU	17
4.1 An Historical Introduction	17
4.2 A Brief History of QMU and How it is Viewed Today	19
4.3 QMU: Management Process or Scientific Process?	25
4.4 Further Observations	27
5 HOW WELL IS QMU WORKING?	29
6 THE FUTURE OF QMU	31
7 REFERENCE	35

Abstract

Quantification of Margins and Uncertainties is a formalism for dealing with the reliability of complex technical systems, and the confidence which can be placed in estimates of that reliability. We are specifically concerned with its application to the performance and safety of the nuclear stockpile, because the test moratorium precludes direct experimental verification. We define performance “gates”, margins and uncertainties and discuss how QMU differs from conventional error propagation. Finally, we review the history of QMU and its meaning, and explore how it may evolve to meet future needs.

1 EXECUTIVE SUMMARY

The national laboratories have begun to develop a formal and systematic scientific framework for the assessment and certification of the performance and safety of the nuclear stockpile in the absence of testing. This approach is called Quantification of Margins and Uncertainties or QMU. The NNSA asked JASON to review the QMU approach in Summer 2004. This executive summary presents our conclusions and recommendations. The main report (JSR-04-330 (U)) provides extensive discussion and support, while a short classified appendix briefly discusses the application of QMU to NNSA problems.

Our study group evaluated the QMU methodology. We addressed both margins and uncertainties. For uncertainties we considered how they are identified, quantified, and used. We focused on the assessment of the nuclear explosive package, and we considered the many possible sources for these uncertainties, including our understanding of the underlying physics, experimental uncertainties, uncertainties inherent in simulations, as well as statistical uncertainty and nondeterministic physics.

1.1 Observations, Conclusions and Recommendations

- The QMU methodology, while currently unfinished, is highly promising. With further careful development, sharpening and fostering, it can become a sound and effective approach for identifying important failure modes, avoiding performance cliffs, and increasing reliability of U.S. nuclear weapons.
- QMU is becoming an effective language for people within the weapons labs to communicate with each other, although the scientific toolbox

which best implements it is still incomplete and evolving.

- QMU can be an effective way of summarizing the output of the weapons labs in assessment and certification of the stockpile to others (NNSA, DOD, STRATCOM).
- Failure modes, gates, and margins form a crucial basis for QMU. The labs are making progress in the initial stages of delineating them, based on expertise and careful, well-planned technical work. Much work remains.
- Margins should be and are being increased in certain cases. This will enhance confidence in the weapons achieving their design performance, and also will improve management of the stockpile.
- Bounding of uncertainties is crucial. There has been substantial progress recently on analyzing, and on enhancing confidence in reducing, some important physical uncertainties.
- In the future, different kinds of uncertainties (*viz.*, physical, modeling, statistical, chaotic, numerical) should be clearly distinguished within the formalism of QMU, and uncertainties should be propagated end-to-end.
- Development of QMU is a joint enterprise of designers, statisticians, physical scientists (both experimental and theoretical), computer scientists, engineers and managers at the labs. They understand that QMU, now in its early stages, represents a culture shift in stockpile stewardship.
- With further development, QMU can become a useful management tool for directing investments in a given weapon system where they will be most effective in increasing confidence in it, as required by the life extension programs. Every weapons system is different, and

comparison of QMU measures *between* weapon systems must be done cautiously. The concentration of resources where margins are most critical must not be allowed to result in loss of key core capabilities that are crucial to the future of the program.

- Margins and uncertainties cannot be studied separately; the uncertainty determines the margin required for confidence, and changes made to increase margins may introduce additional uncertainty.
- QMU is related to project management tools, and their principles and methods should be studied to find lessons relevant to the application of QMU.
- QMU will, and must, allow some resources for basic experimental and theoretical research in order to reduce uncertainties in the future.

1.2 What Must Be Done in the Next Year or Sooner

QMU is now at a critical stage of evolution. For QMU to realize its promise over the next several years, the two nuclear design Laboratories (LANL and LLNL) and the NNSA must take certain important measures immediately, and press them strongly over the next year, and beyond. Here we recommend a list of such urgent measures.

The first two urgent measures are required to clarify and systematize QMU, an ongoing process that must be pursued vigorously:

- It is critical to identify all failure modes and performance gates. The labs need to establish a formal process that identifies each cliff, its location in parameter space and major causes (and probability distributions) for excursions from design performance — transparently, and with the same methodology for all systems.

- The labs need a process to establish interfaces between failure modes and performance gates, over the entire end-to-end sequence of normal or abnormal operation, for all systems in the enduring stockpile.

The next two urgent measures bring QMU to bear on physics uncertainty, and vice versa. The best way to treat an uncertainty is to understand the underlying cause and bound its effects. Ongoing work can and must reduce particular physics uncertainties:

- Recent and ongoing progress in physical understanding has reduced a principal “knob” factor in weapons design. A near term goal should be to reduce this “knob” to less than $\sim 5\%$ uncertainty, over all regimes.
- Pu aging: Develop a database of effects from experiment, and prioritize any actions necessary, using QMU, for each system.

Senior management must take strong leadership in development and use of QMU:

- At present, the meaning and implications of QMU are unclear. The Associate Directors for weapons at the two Laboratories should write a new, and authoritative, paper, defining QMU, and submit it to NNSA, following up the paper by Goodwin & Juzaitis (2003), and building on the current progress noted above.
- Use QMU throughout the annual certification / annual assessment.
- Enter data of UGT failures and excursions into a searchable and accessible database, and mine these data in the context of QMU. Make this part of the training of new designers.

Finally, QMU can help improve the clarity of communications, not only among the labs and NNSA, but also with DoD, Congress, and the press.

2 INTRODUCTION

Since the cessation of nuclear testing in 1991, the nation has evolved new methods for stockpile stewardship. The national laboratories have begun to develop an intellectual framework for the assessment and certification of the performance and safety of the nuclear stockpile in the absence of testing. This approach, the Quantification of Margins and Uncertainties, or QMU, gives the laboratories a common language to articulate their methodology and a means to express a quantitative assessment of confidence in their judgments. QMU may also aid in setting programmatic priorities for stockpile stewardship. The laboratories have reached agreement on many aspects of the framework, and each laboratory is beginning to apply the methodology described by this largely common language to exemplify problems in stockpile stewardship.

In 2004, the NNSA asked JASON to review the QMU methodology, in particular the role of uncertainties. This report gives our conclusions; a classified Appendix provides further details.

We addressed both margins and uncertainties, but we focused on the latter. For uncertainties, we considered how they are identified, quantified, and used. We focused on the assessment of the nuclear explosive package, and we considered the many possible sources for these uncertainties, including our understanding of the underlying physics, experimental uncertainties, uncertainties inherent in simulations, as well as statistical uncertainty and nondeterministic physics.

Specific questions posed by the sponsor were:

- Is the methodology of QMU internally consistent, well defined and broadly understood?

- Is adequate emphasis being placed on the assessment of uncertainties: physical, experimental and numerical? Is there a robust method to identify failure modes and performance gates and the related uncertainties? Do the laboratories have a process to ensure that they have identified a sufficient set of such modes or gates?
- Will QMU improve risk quantification and thereby provide a method of balancing benefits against associated costs? Does QMU provide a useful management tool for prioritizing investments in the Defense Programs scientific program including experimental facilities and their use?
- Does the QMU methodology provide a framework to enable the next generation of trained scientists with no nuclear test experience to make informed decisions regarding the safety, security and reliability of the stockpile?

We are grateful to briefers from the two nuclear weapons laboratories who provided us with abundant, well organized technical information and with insightful discussions. We are especially grateful to Paul Hommert (LANL) and Richard Ward (LLNL) for coordination.

3 QMU PRIMER

Quantification of Margins and Uncertainties (QMU) is a tool for managing science-based stockpile stewardship in a transparent way that facilitates the identification and quantification of failure modes (watch lists), the evaluation of options for responding to changes that occur in nuclear weapons in the stockpile due to aging or other causes, and the setting of priorities for stewardship activities to maintain confidence in the surety, safety, and reliability of the stockpile. It is a formal framework to measure confidence when empirical data are lacking, to identify critical issues and to assist in the optimal allocation of resources in reducing uncertainties. QMU is not itself a procedure for certifying the nuclear weapons in the stockpile, but the formal framework it provides, along with the common language for all parties involved, can be very beneficial to that process. The method of QMU may exploit the existing body of U.S. nuclear weapon nuclear test data, new experimental work, computer-based simulations, and expert judgments. QMU is science-based (but not the same as the “scientific method”). It is also based on long-standing engineering practice.

Margins and uncertainties in the margins are the key ingredients of the QMU approach to stockpile stewardship. Before we can define margin and uncertainty, we must describe in generic terms the essential parts of a nuclear weapon system. A nuclear weapons system can be decomposed into the delivery vehicle and the nuclear warhead(s) it carries. A nuclear warhead can be further decomposed into its nuclear explosive package (sometimes called the physics package) and the supporting engineering components (fusing and firing circuitry, etc.) that are located outside the nuclear explosive package, including the overall container (*e.g.*, reentry vehicle for a missile warhead or environmental shell for a bomb). The nuclear explosive package contains all of the materials that are involved in the fission and fusion processes as well as

the other components (e.g., high explosives) needed to bring about a nuclear explosion from energy released by nuclear fission and (usually) nuclear fusion.

3.1 Gates, Cliffs, Error Propagation and QMU

The fundamental concept in QMU is the *gate*. Before defining a gate, we remind the reader of conventional error propagation formalism. In this formalism a result or output R is a function of one or more input variables: $R = R(x_1, \dots, x_n)$. Then an error or uncertainty δx_i in the i -th input variable produces an uncertainty in R :

$$\delta R_i = \frac{\partial R}{\partial x_i} \delta x_i. \quad (3-1)$$

The effects on R of the uncertainties in the input variables combine. If the uncertainties are mutually independent, a root-mean-square combination is customary (and describes a Gaussian distribution of δR in the limit of many variables):

$$\delta R = \sqrt{\sum_i \left(\frac{\partial R}{\partial x_i} \delta x_i \right)^2}. \quad (3-2)$$

More conservatively, an additive combination may be used:

$$\delta R = \sum_i \left| \frac{\partial R}{\partial x_i} \delta x_i \right|. \quad (3-3)$$

This error propagation formalism is based on a Taylor expansion of R , and is valid when the Taylor series converges rapidly (with one term!). When this condition is met the formalism is a powerful tool. In the Pit Aging study (JSR-05-330) the sensitivities of yield (or other outputs) to variations in physical properties are these partial derivatives.

Gates are defined when this formalism breaks down because at least one of the δx_i , or the corresponding partial derivatives, is large enough that terms in the Taylor series containing it do not converge, or do so only slowly. A practical example occurs in a gasoline engine, where R is the energy released

by burning fuel in a cylinder and x is the energy of the spark. A minimum acceptable output value of R_{min} is specified. A threshold input x_{th} is defined by the condition $R(x_{th}) = R_{min}$.

In general, for x near x_{th} , $R(x)$ is a rapidly varying function; its slope may even be singular. $R(x_{th})$ is not given by a Taylor series expansion about significantly larger values of x , nor is $R(x)$ for x significantly larger than x_{th} given by a Taylor series expansion about x_{th} . This is illustrated in the Figure. In fact, for x close to x_{th} R is likely to be a very sensitive function of other variables (such as temperature, or the richness of the fuel mixture) and may not even be deterministic.

The region of x around x_{th} is referred to as a “cliff” because of the shape of the function $R(x)$. The entire process of ignition is described as a gate because, to good approximation, it only matters whether the gas in a cylinder passes through it ($x \geq x_{th}$) or not ($x < x_{th}$). If ignition occurs, the energy release is almost independent of the pre-ignition conditions. Memory of them is lost, and the result is at least approximately described by a single binary bit—either the fuel ignites and burns, or it does not.

Let x_0 be the nominal value of x which is found if the ignition system is functioning as designed. The manufacturer may specify that x is not less than x_0 by more than a tolerable uncertainty U . Then the margin of successful operation

$$M \equiv x_0 - U - x_{th}, \quad (3-4)$$

(this definition is a convention; others may choose the definition $M \equiv x_0 - x_{th}$). Clearly, $M > 0$ is required for proper functioning of the engine.

For simplicity we have only considered downward deviations in x but, in general, upward deviations must also be considered and will generally be described by different values of $|x_0 - x_{th}|$ and U . Additional uncertainties enter because the function $R(x)$ is imperfectly known and because the choice

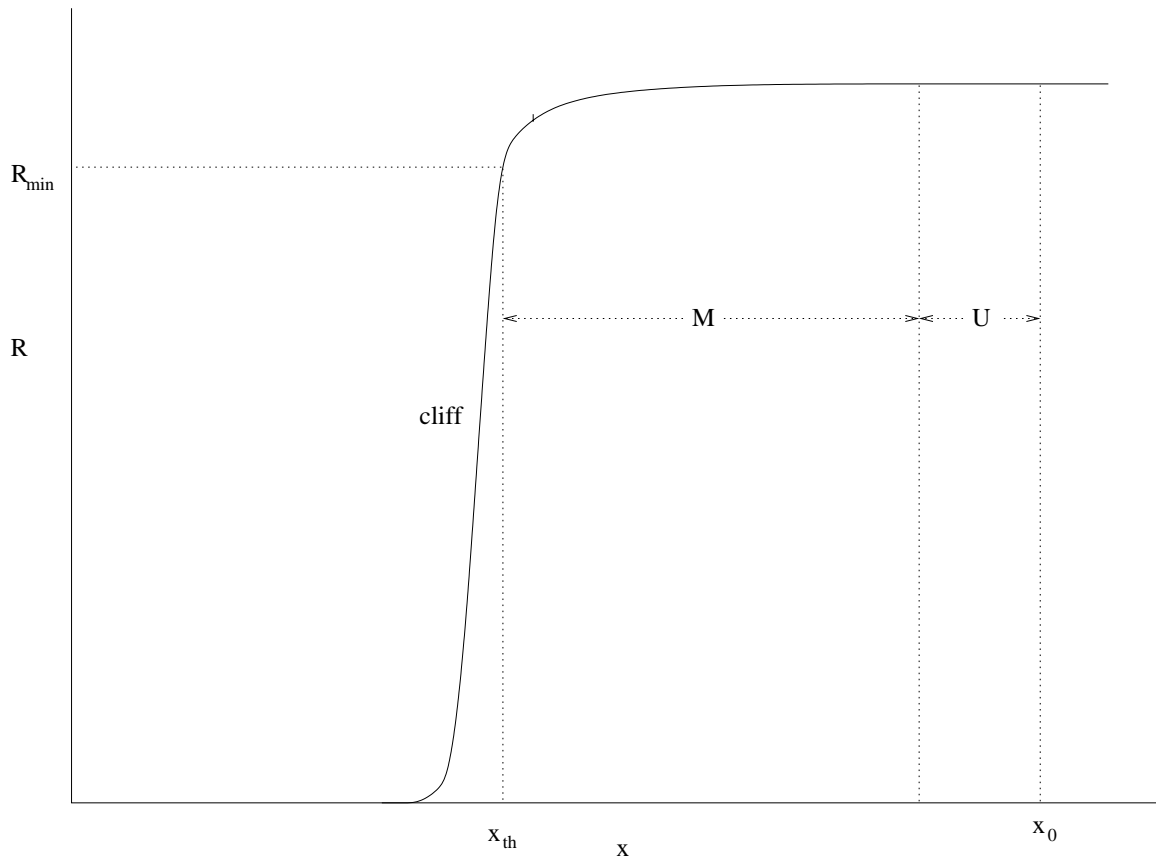


Figure 1: Definition of margin M and uncertainty U showing cliff. x_0 is the nominal value of x and the threshold x_{th} is determined by the minimum acceptable R_{min} .

of a minimum acceptable output R_{min} is necessarily somewhat arbitrary. These extra uncertainties contribute to the effective value of U and reduce M , but may be treated in the same way as the uncertainty in the value of x .

How large must M/U be? The answer depends partly on how reliable the engine must be (a helicopter demands much greater reliability than an automobile). It also depends on the statistics of the distribution of $x - x_0$. If U is an absolute bound on the magnitude of this quantity then $M/U > 0$ (positive margin) is sufficient. If U is the standard deviation of a Gaussian distribution then $M/U > 2$ provides greater than 99.8% assurance of successful operation (more than is required for a weapons system, but probably insufficient for an internal combustion engine). However, if the distribution of $x - x_0$ has long tails (a Lorentzian distribution, for example) then $M/U \gg 1$ is required.

In the real world, parameter distributions rarely have simple analytic forms. For the nuclear components of weapons (in contrast to their non-nuclear components) there aren't enough data to permit their empirical determination, and they are estimated from a combination of numerical simulation and expert judgment. Long tails in the distribution of $x - x_0$ are described, in everyday life, as the chance that something is "broken"; in some cases (such as flight-critical aircraft systems) a combination of conservative engineering, long experience and redundancy have reduced them to extremely low levels.

The quantification process in QMU is intended to accomplish just this, not by collapsing the nuances of expert judgment and the vast experience on which it is typically based into a single number to be placed on a check-list, but by establishing a process whereby estimates of margins, their uncertainties and the reliability of both can be systematically derived. The objective is to be able to re-evaluate continuously the margins and uncertainties in order to assess better their reliability as new information becomes available

(*e.g.*, new experimental data, new simulation results or, most significantly, any new failure modes that might be uncovered).

Uncertainties are more difficult to quantify than the margins themselves, and the reliability of each uncertainty estimate is yet more difficult to establish. Nevertheless, it is generally possible to estimate whether an uncertainty corresponds more closely to a 65%, 90% or 99% confidence level based on a combination of measurement and simulation. It is essential to do this, of course, in order to be able to propagate and combine uncertainties as one attempts to characterize the entire system. Implicit in the process is the requirement to define the underlying statistical distribution, even if only approximately (*e.g.*, normal *vs.* log-normal). The end result ought to be a process for estimating confidence levels, and a clear language for describing the uncertainties associated with key margins.

3.2 Uncertainties

There are several different kinds of contributions to uncertainty in the performance of complex physical systems:

1. Uncertainties in physical properties—in principle, these are reducible by experiments.
2. Uncertainties in physical processes responsible for observed behavior—in principle, these are reducible by experiments.
3. Uncertainties in physical realizations—in principle, these are reducible by tightening manufacturing tolerances and control.
4. Uncertainties in the environment—in principle, these are controllable.
5. Uncertainties resulting from the non-deterministic nature of some complex physical processes.

3.2.1 Indeterminacy (Chaos)

The last item is of particular interest. A familiar example of a non-deterministic physical system is the weather. At the same time of year, at the same time of day and at a specified location the temperature, humidity and rate of precipitation vary substantially from one year to the next, or from one day to the next, despite essentially identical forcing functions (geography, insolation, *etc.*). Similarly, climate is substantially non-deterministic, with warmer and cooler periods, ice ages and interglacials, wetter and drier epochs, *etc.* So is the flapping of a flag in the wind and many other turbulent flows. Of course, all these systems do respond in a predictable manner to variations in forcing functions (winter is cooler than summer, for example), but to this must be added a significant non-deterministic component.

Unfortunately, we do not know to what extent the performance of nuclear weapons is non-deterministic. There are no homogeneous series of tests, and the occasional significant deviation from expectation is as plausibly the result of a deterministic process as of indeterminacy. On the other hand, indeterminacy is important, because it may pose the ultimate limit to reliability.

Can we learn anything about the possible presence of indeterminacy in nuclear weapons? Studies of indeterminacy in physical systems began with the work of the mathematician Poincaré a century ago, and much of our present understanding dates from that period. Interest was revived when Lorenz stumbled over indeterminacy in a weather calculation in the 1960's. He restarted a computer simulation with parameters copied from a print-out (and therefore differing in low-order bits from their values stored in the computer) and discovered that these small differences in input made large differences in the results later in the simulation (a quantitative theory of such phenomena in differential equations had earlier been developed, without ref-

erence to computer integration, by Liapunov). This is often described as the “butterfly effect”—the beating of a butterfly’s wings will change the weather a year hence. Of course, the butterfly effect is only a hypothesis—we do not have two weather systems, one with and one without the butterfly, with which to test it, but its reality is generally accepted.

In the field of nuclear weapons, analogous numerical experiments can be performed, introducing small perturbations to see if they produce larger consequences. This may be significant where there are hydrodynamic instabilities. If indeterministic processes occur in such areas as sub-grid turbulence, then their consequences are not directly accessible to system-scale numerical experiment, but could be investigated with subscale calculations. The results of these subscale calculations could then be coupled to system-scale calculations. Indeterministic processes on larger scales, such as longer wave hydrodynamic instabilities or in chemical reaction networks, *etc.*, could be studied directly with system-scale calculations.

3.2.2 Human uncertainties

In addition to the physical uncertainties that experts are able to estimate, there are uncertainties that have to do with the competence and credibility of the experts. When a non-expert citizen or member of Congress is judging the reliability of the stockpile, uncertainties of the second kind are at least as important as uncertainties of the first kind. The member of Congress may reasonably say to the expert, “Why should we believe what you say?” History is full of disasters which the experts of the day had declared to be unlikely or impossible. To take a recent example, NASA at one time claimed that the chance of a fatal crash of the Space Shuttle was one in a hundred thousand. The two Shuttle crashes proved that the experts failed by choosing to ignore risks whose existence (though not their quantitative magnitude) were known—these were not “unknown unknowns”. Members of

Congress (and their staffs) will try to decide whether the experts in NNSA are more reliable than the experts in NASA. In making a judgment of the overall reliability of the stockpile, uncertainties of expert judgment must be taken fully into account.

Fortunately, we have a massive historical record that gives us an objective measure of the reliability of the nuclear experts. The record consists of the results of the nearly 1200 nuclear tests conducted by the USA since 1945. The tests were of many kinds, conducted for a variety of purposes, but they give us, irrespective of technical details, a crude measure of the competence of the experts who planned them and carried them out.

Let N be the total number of tests that confounded the experts, either by failing to give the expected yield or by giving a nuclear yield when they were not expected to. Then the fraction

$$f = (N/1200) \tag{3-5}$$

gives an objective measure of the frequency of expert misjudgements and mistakes. To determine N it is necessary to go through the historical record of tests and pick out those that failed unexpectedly. This we have not done.

Informal discussion with people from the weapons laboratories leads us to estimate that N is about 24 and f is about 0.02.¹ A historian who is not a weapons-designer or an employee of NNSA could be asked to make a careful determination of N , and the resulting value of f then be made public as an objective measure of uncertainties resulting from limitations of expert judgment.

If the same methodology had been applied to estimate the reliability of the Space Shuttle, based on the historical record of space launches, the resulting value of f might have been as much as 0.10. That would have over-

¹Curiously, this is about the same as the rate of catastrophic Shuttle failures. However, the Shuttle is (was) operated to minimize risks, while some nuclear tests were deliberately performed to investigate regions of uncertainty.

estimated the probability of a shuttle crash, which empirically has been in the range 0.01–0.02, but not by a large factor. In the case of the nuclear stockpile, the value of f probably also over-estimates the probability of human errors affecting weapon performance, because the modern stockpile weapons have been tested and analyzed more carefully than earlier generations of weapons. But the member of Congress need not assume that f is an over-estimate. If the value of f is around 0.02, this should be an acceptable level of uncertainty for the stockpile weapons. It is smaller than the uncertainty of the command-and-control and delivery systems that are required for the operation of the weapons.

It is politically important that uncertainties of this kind should be included in any public statement of stock-pile reliability because the credibility of the statement depends on the credibility of the experts.

4 AN HISTORICAL VIEW OF QMU

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?

—T. S. Eliot, “The Rock”

4.1 An Historical Introduction

During the JASON briefings on QMU, it became evident that there were a variety of differing and sometimes diverging views of what QMU really is and how it is working in practice. To some extent, NNSA headquarters staff, lab management, and lab scientists and engineers differ with each other and among themselves. Confusion among these people will surely lead to confusion and distrust among people who have critical concerns and responsibilities for nuclear weapons, including DoD customers, the Secretary of Energy, and others, who are told that something termed QMU is essential to stockpile stewardship. In what follows we attempt to gather together these views as we heard them at the briefings and as they are found in several documents dating from 1997 to the present. We then attempt a definition and a program for QMU which could reconcile the many beliefs now existing. It may happen that no one else agrees with our views. But we emphasize the importance of finding a universal definition of QMU and its implications which can be explicitly accepted and implemented by NNSA, lab management, and lab technologists.

We note at the outset that the QMU problems faced by the design labs often differ from those faced at Sandia, where full-scale component tests

in sufficient numbers to allow statistical analysis are usually possible. The characteristic feature of QMU at the design labs is that the underground test (UGT) data base is simply not broad enough or deep enough to allow confident statistical analysis, and could not become so even if UGT were resumed. Although there were about 1000 US UGTs, there is often little breadth in covering a specific weapons function (whether successful or anomalous) in a specific weapons type. And UGTs did not go deep into weapons function in many cases; they were perforce instrumented mostly for global and integral quantities, such as primary and secondary yields and a few other markers of weapons performance. Our report covers what QMU should be for the design labs, given that UGTs are unlikely to be resumed in the near future and that the UGT database may well contain less information than is needed for stockpile maintenance today. This is not to slight the importance of finding a definition and process for QMU that all three labs will agree with and apply, but our comments below refer to the design labs' problems only. [We received no briefings from Sandia, although some Sandia people attended.]

Following Eliot, we distinguish *information* from *knowledge* from *wisdom*. For our purposes, information is what scientists and engineers gather through experimentation, large-scale simulations, application of statistical methods, and the like. Knowledge is gained by organizing information; it includes analyses of various sorts. And wisdom (according to the American Heritage Dictionary) is "The ability to discern or judge what is true, right or lasting; insight". In the context of stockpile stewardship, wisdom means making the right decisions in managing stewardship, in the certification process, and in conveying to the US government and people a strong confidence in these decisions. As Eliot suggests, one can slight knowledge if the process of information-gathering is dominant, and one can slight wisdom if mere organization of information is dominant. We believe that the history of stewardship since the end of UGTs in 1991 reveals a tendency to confuse these three concepts of Eliot's, along with several efforts on the part of the

weapons labs and headquarters management to end this confusion. QMU is now the watchword, in the sense that all stewardship players agree that there must be something like it that drives stewardship. But exactly how can it unite information, knowledge, and wisdom?

In fact, the word “wisdom” is not commonly used to describe the endpoint of QMU. Two words sometimes used interchangeably and confusingly to describe this endpoint are confidence and reliability. These are not at all the same concepts. In our view, confidence is what the Secretary of Energy and the President need from the weapons labs; it is not a number like M/U, nor is it the summary of a statistical analysis. Reliability, on the other hand, may well be quantified by various scientific and statistical analyses. Reliability may be the concept used by the labs to express certain of their QMU findings quantitatively, and as part of their own process of building confidence internally, but confidence is what they must pass on to higher authority outside NNSA.

First we review some of the steps taken and views held, at times ranging from several years ago up to this year (2004)’s JASON Summer Study, that led to the concept of QMU.

4.2 A Brief History of QMU and How it is Viewed Today

To understand what QMU ought to be, one should understand the needs that called QMU into being. After UGTs ceased in 1991 and Vic Reis (then Assistant Secretary of DOE responsible for Defense Programs) in 1993 inaugurated science-based stockpile stewardship (SBSS), the design labs went through a frustrating period in which it was difficult for them to see how they could maintain confidence in the nuclear stockpile with no UGTs as the years went by, no matter how much science and engineering they did. Much

of the science and technology (S&T) effort of those early post-UGT days seemed to outsiders to be largely unrelated to pressing and specific issues related to maintaining the stockpile. The design labs' S&T efforts were seen by many as a giant sandbox of unfocused level-of-effort work, unconnected to budgetary and other priorities. In short, science-based stewardship as practiced by the labs was seen by many as a process of gathering information, with no well-defined path to knowledge and wisdom. To be sure, there were programs that dealt directly with specific stockpile issues, such as significant finding investigations (SFIs), but no clear relation between these and the rest of SBSS.

There were several views on this state of affairs: One was that post-UGT information-gathering could never succeed, and that UGTs must resume. Another group believed that neither UGT nor SBSS was needed; it would suffice to maintain the stockpile by rebuilding it, as needed, precisely as each weapon had been originally manufactured. This view is now muted, given the evident inability (for environmental and other reasons) of re-manufacturing the stockpile to legacy specifications. Some (originally few, later many) believed that ultimately SBSS would lead to real confidence in certification. In part because of the promise of future experiments at NIF (the National Ignition Facility), the spectacular growth in simulation capability, advances in above-ground experiments (AGEX), such as hydrodynamic experiments and sub-critical experiments, and enhanced surveillance including aging studies, the last view has come to prevail. But even so, just how to manage SBSS was very unclear.

A suggestion which anticipated QMU, and played some role in its furthering at Livermore, occurs in the 1997 report of the National Security Advisory Committee at Livermore (one of the JASONs participating in the QMU Summer Study was a member of that committee). It reads:

The quality of science in isolated stewardship projects is outstanding. A weakness of the S&T program, however, is the lack of a coherent strategy. . . We suggest construction on an ongoing basis of a “Blue Book” set of documents which: a) quantifies the impact of various S&T projects on their direct relevance to stockpile problems; b) sets priorities for which S&T projects and at what level of precision are the most important; and c) provides for necessary peer review, independent analyses, and cooperative interlab efforts. Despite the difficulty of developing these Blue Books, we feel that their existence, even in the early years when they will be very incomplete, will be of great help in defending the program.

In 1999, the Campaigns process was inaugurated. It constituted a set of (now) 17 programmatic efforts to direct critical science and engineering needed for certification, and was separate from another set of programs termed Directed Stockpile Work (DSW). DSW, as its name implies, involves activities on specific weapons and components as needed to solve problems, current or anticipated. There were also separate program activities in facilities construction and infrastructure. While Campaigns did bring some organization to the weapons labs’ S&T programs, it did little to relate S&T work to DSW issues. Campaigns are largely a process of organizing information and turning it into knowledge, although they does not capture all the knowledge needed to generate wisdom.

In 2001-2002 NNSA became convinced that it needed a new way of organizing and managing stewardship, at the same time that the weapons labs were beginning to see common grounds in the new and undefined process called QMU. Two specific outcomes of this struggle to capture the essence of QMU are 1) a 2003 white paper (hereafter GJ) by Bruce Goodwin, Associate Director for nuclear weapons at LLNL, and Ray Juzaitis, at the time Asso-

ciate Director for nuclear weapons at LANL, and 2) a so-called “Friendly Review” of QMU as practiced at the labs. The Friendly Review was done in late 2003 and early 2004 by the Defense Programs Science Council at NNSA Headquarters. Neither GJ nor the Friendly Review was intended to be a definitive statement.

The process envisaged in GJ, ultimately leading to certification, gives a central role to a watch list of possible and credible failure modes of specific weapons. Items get on the watch list if their function is critical to nuclear weapons performance. The goal is to quantify both the margins and the uncertainties associated with these watch list items. Confidence (and thus wisdom) arises when margins are demonstrably and sufficiently larger than uncertainties. The watch list is derived from a time line of critical events in a nuclear explosion.

In the context of a gasoline engine, some of these events would be the opening and closing of valves, the firing of the sparkplugs, the ignition of the fuel in each cylinder, the meshing of the transmission gears, *etc.* For each event one or more performance gates are defined, representing the acceptable range of values for specified physical variables based on available margins (M) and uncertainties (U) for these variables. For example, for the ignition of the fuel the relevant variables might be the state (usually parametrized by its internal resistance) of the battery, corrosion or fouling of the sparkplug, fuel mix in the cylinder, timing of the distributor, *etc.* Each event has potential failure modes, identified in part by the performance gate analyses. Proposed values for M and U are given to critical components and functions.

The Friendly Review was not intended to be a formal review of the QMU process, but simply a quick and informal way for NNSA HQ to find out what lab people thought QMU was and how they planned to implement it. It is evident from the Friendly Review and from various lab reactions to it that there has been a certain lack of coherence between lab and headquarters

views of QMU, although both sides have declared their enthusiastic allegiance to it. Our remarks on the Friendly Review are not intended as criticism either of the reviewers or of those reviewed, but simply to suggest that the QMU process is indeed inchoate now and that all involved must renew their attempts to define and institute the QMU process.

In very brief summary, the Friendly Review says that QMU is a major step in the right direction. QMU provides transparency and common language to enhance weapons assessment and certification, and provides a science-based methodology to do this. The Friendly Review concludes that QMU will provide significant benefits to the weapons programs and NNSA, including a systematic way to anticipate and identify failures and the potential for cost savings. We agree with these findings. The Friendly Review also notes that QMU is in its earliest stages, and that there are some areas of significant concern and major challenges yet to be met. It says that QMU is not uniformly articulated and not well-understood or followed by most designers (indeed, it is claimed that there is a certain amount of hostility to QMU in this community); that the labs are not working together on implementing QMU; and that QMU has made obvious the “unfortunate” isolation of the design community from outside science. Further, it says that the labs have not shown that they have a method, even in principle, to calculate the overall probability that a weapons type will perform as designed, and that lab efforts in QMU are compartmented and disorganized.

While we are not aware of any formal response from the labs to the Friendly Review, we have heard informal comments from lab management expressing concern that the Friendly Review was, in certain cases, misinformed and misguided. Some have said that there is in fact a commitment to QMU at the labs, rather than the hostility perceived in the Friendly Review. Concern has been expressed over the perception that the Friendly Review takes it as given that QMU is a form of statistical analysis, while in fact the UGT data base is not sufficiently rich for traditional statistics applications.

And lab people have pointed out that the Friendly Review has little to say about the role of margins, which is central to the QMU process.

Next we turn to very recent history: the reactions of people involved in putting together QMU as we heard them during informal sessions at our Summer Study and elsewhere. We heard reactions from NNSA HQ people; weapons labs management through the Associate Director level; and designers, engineers, statisticians, and lab science staff.

In general, management, whether at NNSA HQ or at the labs, sees QMU as a management process, designed to make transparent the ways in which lab science informs decision makers and to furnish a framework for prioritizing and organizing the science and technology used for stewardship. Some spoke of changing a lab science and technology culture where science was done for its own sake, without regard to its necessity, sufficiency, and urgency in specific stockpile maintenance programs.

Lab scientists, designers, and engineers felt that there was indeed a culture change among them underway, but they were not sure what it was; there seemed to be many versions of QMU. Some saw statistical analysis as fundamental to QMU; others did not. Some mourned what they felt would be a withering away of the role of expert judgment once QMU became established. Others thought that expert judgment would have to be an integral part of the QMU process. Some senior designers saw a culture change toward simulations and away from understanding UGT data among younger designers, and warned against QMU as ultimately a simulations-driven tool. Some younger designers worried that QMU would drive them into an endless search for impossibly quantitative answers to ill-posed questions.

Based on what we have reviewed above, we draw one conclusion:

There is general agreement that QMU can be an effective way of summarizing the output of the weapons labs

in assessment and certification of the stockpile, and in conveying confidence in the reliability of the stockpile to others. There is also agreement that QMU can be a worthwhile management tool, useful for prioritizing and organizing stockpile-stewardship science and technology into a technically and cost effective set of programs, as well as for identifying failure modes and actions needed to avoid them, including margin enhancement where appropriate. There is no general agreement on what QMU means to the various scientific processes commonly used in science and technology, or whether QMU is in some sense a new such process.

In T. S. Eliot's terms, there seems to be a consensus that QMU has something to do with knowledge and wisdom, but there is no consensus on what it has to do with information.

Next we will draw some parallels between QMU and certain other management tools.

4.3 QMU: Management Process or Scientific Process?

The managers of large construction and engineering projects (including those at the labs, such as NIF, DARHT, and LANSCE) obviously need tools for overseeing their projects. We do not refer here to standard budgets and schedules, but rather tools for making sure that the project will be the best it can be, while meeting performance, cost, and schedule goals. Elaborate commercial software is available for this sort of thing; for example, Primavera, which was used at LANSCE. Other kinds of analysis are routinely performed for such large projects. An example is RAMI (reliability, availability, maintainability, inspectability), which has a number of formulas for combining

what is thought to be known about components, such as mean time between failures (MTBF). The output of the RAMI process is supposed to provide knowledge about the four components of RAMI. These project planning and management tools do not attempt to tell designers of lasers or accelerators (for example) how to design these things, but do ask for quantitative information from designers as input to the planning and management tools. And if the answers give cause for concern in the planning tools (for example, a laser or linac section has too short a MTBF, or has a cost overrun), management can interact with the designers to fix the problem. Designers and engineers in these large projects are aware, and accept, that they must be able to furnish sensible and defensible answers to the questions asked by the management tools.

Although we certainly do not suggest a direct comparison between QMU and RAMI or other large-project tools, we do suggest that NNSA and the labs should look at QMU in a similar spirit. We conclude:

QMU is a management tool, intended to couple SBSS science and technology to critical priority issues for specific stockpile systems, including potential failure modes and ways to enhance margins when needed. [At the same time, NNSA and the labs must explicitly recognize and reward a reasonable amount of independent research not directed toward specific stockpile issues.] Another major value of QMU must be in organizing the results of lab science and technology in terms that make the tasks of stockpile assessment and certification transparent. QMU's terms of reference, margins and uncertainties, must underpin the elaboration of goals and directions for science and technology (such as watch lists, performance gates, and the like), but must allow for flex-

ibility in techniques and for the role of expert judgment in reaching these goals.

We recommend that NNSA and the labs study the principles and methods underlying management tools for large construction and engineering projects, to see if QMU can be realized in a similar spirit.

4.4 Further Observations

1. In the briefings we heard, there was substantially more time devoted to understanding and ultimately decreasing uncertainties than there was to increasing margins, and there seemed to be a tacit assumption that margins and uncertainties could be understood separately. In fact, they should be studied together, for at least two reasons. First is the obvious reason that if one finally reaches an irreducible level of uncertainty, attention must be paid to the corresponding margin; if it is insufficient, steps may have to be taken to increase it. Second, steps taken to increase a margin may also act to increase one or more uncertainties, depending on the extent to which a particular uncertainty can be quantified without resort to UGT. It will be important to understand this sort of correlation between margins and uncertainties; to assume they are uncorrelated may well be wrong in critical instances. In other instances, there may be little correlation.
2. It is perhaps natural to believe that any system making use of uncertainties must ultimately be based on statistics. But at least in the special case of stockpile stewardship, we think this is, at best, an oversimplification. The data base for UGT is simply too small and not sufficiently diverse (for specific weapons functions) for statistics to be

relevant. While it is appropriate to apply statistical analyses to large numbers of numerical experiments, one must still be cautious. As the simulation parameters encroach on the regions of cliffs, where the output of some module begins to be insufficient to drive the next module in the chain to satisfactory function, numerical simulations begin to break down and statistical analysis might add little value. Statistical analysis may well have an important role in understanding the science and technology output that QMU makes use of, but it should not be considered as the central theme of QMU itself.

3. The study of such cliffs, which relate one weapons function to another, is an important tool in relating margins to uncertainties that complements statistical analysis. The UGT database is limited, but far from empty, in data related to cliffs. Briefers pointed out that a very good indicator of cliffs in numerical simulations is the incipient chaotic behavior of the simulations as parameters defining cliffs are varied.
4. Neither QMU nor any other proposed approach to stewardship should become stalled by the issue of “unknown unknowns”. Invoking unknown unknowns is a way of stopping anything short of resuming UGTs. By this time, there is little in the collective knowledge base of all designers that is not understood qualitatively in the design of nuclear weapons. The trouble may come when the older designers retire and their knowledge is not fully passed on to the next generation. Fortunately, many senior and retired designers are still available in the weapons labs. The labs do their best to transfer knowledge from the older generations to the newer. One very important part of this transfer, well-understood by lab scientists, is to construct watch lists that are as complete as possible and in other ways to incorporate the best expert judgment needed for QMU.

5 HOW WELL IS QMU WORKING?

The success of QMU depends upon how the scientists and engineers whose jobs include assessment of the expected performance of components and subsystems of the nuclear explosive package make use of it. It also depends on how the managers who must make the assessment of the total weapons package use it to determine if they can “certify” the weapon for continued inclusion in the stockpile. One measure of the success of QMU is that the people in the first category use QMU in a way that makes the managers more confident of their ability to certify weapons and better able to decide how to allocate limited resources in the stockpile stewardship program. In fact, it is also clarifying the assessors’ jobs. They have a metric, M/U, rather than just “expert” judgment, to decide if a subsystem or component will work, but they still have the freedom to decide when a particular value of this metric is sufficient to assure confidence in a particular item on their list. The assessors seem to be satisfied with this situation.

Given this starting point, we can ask how well does QMU appear to be working when applied to specific weapons. We heard extensive briefings on several systems. Many potential failure modes have been identified and studied, including well-known ones, such as the cliff in total weapon yield *vs.* primary yield. For most of those failure modes, the nominal performance of the component in question gave a margin for success that was large compared to the uncertainty in the performance, giving an M/U that was in the range 2–10. However, for at least one component of one system, a problem was identified with an unsatisfactory M/U and a design change was introduced that will (when fully implemented) increase the M/U for that system element to a high value, at the expense of an acceptably small decrease in the M/U of another system element.

This was an example in which resources were concentrated on a par-

ticular problem because the QMU procedure indicated that other system elements were unlikely to fail. Notice that this process does not eliminate the possibility that we have missed a failure mode entirely (the familiar “unknown unknowns” concern), but it does point out which of those that are identified need resources applied to their elimination or improvement.

6 THE FUTURE OF QMU

As the stockpile ages, but our body of information obtained from a variety of experiments short of nuclear tests increases, the universe of knowledge to which QMU methods can be applied will expand. “Ground truth” will not be limited to archived weapon test data as the labs try to reduce uncertainties. A suitably designed subcrit series, or series of experiments on the Z-machine or the Omega laser or a series of gas gun experiments can be fielded and compared with appropriate models to differentiate among them.

On the other hand, waiting for the NIF to benchmark weapon physics models is not a good idea. NIF will be too busy trying to achieve ignition by 2010 to do much serious weapon physics until ignition is achieved. Once ignition is achieved it may be very useful, but that is far off. The further we get from tested manufacturing processes and materials, the more important QMU will be to establish confidence in stockpile reliability.

If the life extension program should change to a rebuild program a QMU analysis will be necessary to obtain confidence in the reliability of a weapon which is not an “exact” copy of one which has been tested. Some day, weapons may be redesigned to increase margins. Weight constraints will then be relaxed (because the degree of MIRVing will be reduced), and their number will be far below that deployed during the Cold War. Then designing for manufacturability and ease of maintenance could reduce the cost of the production facility.

We have so far considered each gate an independent entity. The trajectory of a given weapon system through a succession of performance gates was treated as a Boolean AND operation, with all knowledge of the system trajectory lost between gates. While this is a sensible initial approach, in some cases an end-to-end modeling approach in which the parameters at

one gate (where in the gate it passes) are propagated to the next gate may be necessary. Of course, this negates the idea of a gate. The question is the validity of the gate concept. Initially, it is attractive—an engine cylinder either fires or it doesn't—but the requires quantitative investigation, empirical when possible, and by simulation when not.

As it becomes established and therefore accepted, QMU naturally leads to a prioritization of risks associated with the weapon system functioning as intended. Processes having large margins relative to uncertainties can be considered reliable, and therefore require less direct attention in the Stewardship Program. This offers a basis for deciding what issues need greatest attention, and therefore what activities need to be pursued with highest priority. The intellectual resources associated with designing and maintaining the weapon systems are stretched thin, as the relevant expertise takes years to develop and the community of experts is relatively small. Therefore, it is important that efforts be focused on the most significant issues, with minimal attention distracted by lower-risk aspects of the weapon's full functionality (reliability, safety and security). QMU can play a key role in identifying the relevant priorities.

There are at least three ways in which the temptation to apply QMU directly to budget-prioritization must be avoided, however. First, it is important to maintain enough duplication of expertise and effort to sustain quality. That is, there must be enough redundancy to allow full peer review of each activity associated with LEPs, remanufacture and certification, and the underlying science and engineering research. Thus, QMU can help define immediate priorities, but there needs to be adequate capability to ensure quality for all levels of effort including those judged, at any given time, to be of lower priority.

The reason for this last comment is elaborated in the second point, which is that there must be adequate expertise available to identify and respond to

new problems or unexpected failure modes. In other words, the success of QMU should not be that it can establish a rigid set of priorities, but should be based on the ability of QMU to respond to and prioritize newly discovered problems (whether new problems, caused by aging for example, or existing problems that had not previously been identified or fully appreciated). To be able to reliably identify as high priority an issue previously considered to be less important is a fundamental requirement of QMU.

Finally, prioritization of efforts has to be modulated by the need to maintain expertise across the entire weapon system, and its processes. That is, a baseline of effort needs to be maintained across all activities, including those judged as being of lower priority. Of course, less effort should be put into lower-priority activities (i.e., those bearing on processes with higher margins relative to uncertainties), but there needs to be enough ongoing activity even regarding "reliable" (high-margin) processes in order to maintain expertise and to allow for the possibility of revising previous estimates of reliability (and responding to those revisions).

7 REFERENCE

Goodwin, B. T., and Juzaitis, R. J. National Certification Methodology for the Nuclear Weapon Stockpile (U) Draft working paper, March, 2003.

DISTRIBUTION LIST

Assistant Secretary of the Navy
(Research, Development & Acquisition)
1000 Navy Pentagon
Washington, DC 20350-1000

Principal Deputy for Military Application [10]
Defense Programs, DP-12
National Nuclear Security Administration
U.S. Department of Energy
1000 Independence Avenue, SW
Washington, DC 20585

Strategic Systems Program
Nebraska Avenue Complex
287 Somers Court
Suite 10041
Washington, DC 20393-5446

Headquarters Air Force XON
4A870 1480 Air Force Pentagon
Washington, DC 20330-1480

Defense Threat Reduction Agency [6]
Attn: Dr. Arthur T. Hopkins
8725 John J. Kingman Rd
Mail Stop 6201
Fort Belvoir, VA 22060-6201

JASON Library [5]
The MITRE Corporation
3550 General Atomics Court
Building 29
San Diego, California 92121-1122

Ambassador Linton F. Brooks
Under Secretary for Nuclear Security/
Administrator for Nuclear Security
1000 Independence Avenue, SW
NA-1, Room 7A-049
Washington, DC 20585

Dr. Martin C. Faga
President and Chief Exec Officer
The MITRE Corporation
Mail Stop N640
7515 Colshire Drive
McLean, VA 22102

Dr. Paris Genalis
Deputy Director
OUSD(A&T)/S&TS/NW
The Pentagon, Room 3D1048
Washington, DC 20301

Brigadier General Ronald Haeckel
U.S. Dept of Energy
National Nuclear Security Administration
1000 Independence Avenue, SW
NA-10 FORS Bldg
Washington, DC 20585

Dr. Theodore Hardebeck
STRATCOM/J5B
Offutt AFB, NE 68113

Dr. Robert G. Henderson
Director, JASON Program Office
The MITRE Corporation
7515 Colshire Drive
Mail Stop T130
McLean, VA 22102

Dr. Charles J. Holland
Deputy Under Secretary of Defense Science
& Technology
3040 Defense Pentagon
Washington, DC 20301-3040

Records Resource
The MITRE Corporation
Mail Stop D460
202 Burlington Road, Rte 62
Bedford, MA 01730-1420

Dr. Ronald M. Sega
DDR&E
3030 Defense Pentagon, Room 3E101
Washington, DC 20301-3030

Dr. Alan R. Shaffer
Office of the Defense Research and Engineering
Director, Plans and Program
3040 Defense Pentagon, Room 3D108
Washington, DC 20301-3040