LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-653196

Applying QMU to Nuclear Weapon Stewardship

P. J. Raboin

April 15, 2014

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Applying QMU to Nuclear Weapon Stewardship

Abstract

Quantification of Margins and Uncertainty (QMU) is a principle means and metric by which nuclear weapons performance is measured [1, 2]. At Lawrence Livermore National Laboratory (LLNL), Nuclear weapon QMU is predominately a Physics assessment of the performance margin and associated uncertainties of not achieving the designed nuclear performance. To more completely apply QMU to Nuclear Weapon Stewardship, a Systems Engineering or systematic approach is being employed to identify and define requirements, allowing assessment of performance with QMU. This paper describes the basics of QMU and how it is applied below the highest level nuclear performance function down and throughout the nuclear weapon design architecture. The intended audience for this paper is a novice QMU practitioner at LLNL.

The application of QMU to the entire set of nuclear weapon functions demands a broad engagement of Science and Engineering disciplines. Using a hierarchical flow-down requirements structure, performance functions and failure modes are identified for all levels of system requirements. There can be multiple failure modes per functional requirement. For each function there is a spectrum of possible outputs that can vary from outright failure to excellent performance. The possible consequences and impacts on the next higher level functions will also vary. Since performance variations exist in a weapon functional hierarchy, the assessment process needs and benefits from a system perspective. Incorporating QMU into assessments using a Systems Engineering approach to stockpile management provides a means to quantify weapon performance risks against a multitude of possible failure modes and from the highest level functions down to the lowest level component functionality.

Performance risk management is a core practice of nuclear weapon management. Assessing the likelihood and consequences of the many possible failure modes is done yearly as part of Annual Assessment Reviews (AAR). When annual surveillance discovers an anomaly defect as part of weapon surveillance, investigations are launched and QMU is a key tool for quantifying performance impacts. QMU assessments are separated into two pieces, failure mode assessments and QMU analysis

Background

While the preponderance of QMU assessment work is classified, the basics of QMU are unclassified. This paper describes QMU, explains why it is important and how it fits into the larger scope of stockpile stewardship activities. Appendix A gives a wide range of analytic approaches for estimating confidence levels through statistical analysis, explains Sandia K factors used to select sample sizes appropriate for reliability and confidence level requirements and finally it surveys first and second order reliability analyses and a variety of computational methods. Appendix B covers the statistics used as the basis for nuclear weapon surveillance sampling quantities. The reference papers are available in the LLNL, NWE share folder \\wci-cl2\wci\NWE\QMU.

Nuclear Weapon Stewardship before QMU

The original nuclear weapon stockpile stewards benefitted from two key experiences lacking in modern stockpile stewardship: original design development and full scale production. The original physicists, chemists and engineers gained an understanding of the numerous design parameters that affect weapon performance through the experience of building a large nuclear weapon stockpile, and they

learned to utilize design margins for assured functional performance. This is not to say that they had quantitatively determined the minimum set of performance parameters, but they did gain knowledge of the necessary set of performance parameters. From testing failures, they learned how their designs performed and how to avoid failures with better performance margins.

From the experience of full scale production, the original developers were informed by the validation of production build data to confirm that weapons were built to design specifications. From their direct experience, they also knew when as-built performance parameter variations were large. This production knowledge provided the original stewards with insights into understanding weapon performance risks and the qualitative likelihood of potential failure modes. Weapon designers strove to build the stockpile to the same specifications as was tested in Under Ground (nuclear) Testing (UGT). They quantified the acceptable design specifications. Rarely could they quantify performance parameter specifications that defined the failure cliffs. Performance uncertainty quantification was typically limited to the experimental uncertainties associated with UGT.

There were numerous reasons that QMU was added to nuclear weapon assessments. The first occurrence was for the W87 LEP Certification (around 2000), when QMU provided a quantitative basis to judge LEP changes and improvements. This experience proved that assessments with QMU demanded a higher level of rigor and weapon knowledge than previous certifications. It was also realized that QMU could be the means to train the next generation of nuclear weapon stewards. Through the exercises of quantifying performance margins and uncertainties, physicists, engineers and chemists gain job experiences similar to design development and full scale production. Specifically, current weapon stewards need to learn, through study and quantification, about the numerous design margins built into nuclear weapons. They need to consider the numerous possible failure modes and to quantify the design margins, i.e., how close is the nominal performance to a failure cliff. Just as the original stewards knew how well the weapons had been built, so too should modern stewards examine the quality of the original build data, and to quantify the variation in performance parameters to improve the likelihood assessment of a failure mode risk. To successfully execute a QMU process, the modern designer rediscovers the original designers' intent and explores production records in order to quantify the nominal weapon performance and its variation.

QMU Origins, Motivations and Reporting

The academic bases of QMU are rooted in statistics and engineering reliability or design reliability methods. The following list of textbook references provides instruction on margin over variance ratios and how they relate to the probability of failure [3-11]. The earliest mention [12] of a probabilistic failure approach based on overlapping statistical distributions of loadings and strengths come from a publication by Sir Alfred Pugsley in 1939 [13]. He recognized that statistical variations in the strengths of aircraft when subjected to probabilistic variations in aircraft loadings could result in airframe failures.

Descriptively, the QMU Confidence Factor (CF) is a metric defined as the performance margin divided by the performance uncertainties. The margin basis is dependent on a selected failure mode. The CF has similarities to other metrics with names, such as Safety Index, Safety Margin and Reliability Index. QMU can be related to other design metrics such as the engineering safety factor, the Sandia K-factor and first and second order reliability methods (see Appendix A).

Aviation [14, 15], space vehicles [7, 16, 17], and probabilistic mechanical design employ a variety of statistical methods, safety factors, and reliability/safety indexes to improve and optimize their designs. In particular, air and spacecraft designs use a reliability requirement and confidence limit specification.

They grade this approach based on the consequence of a component failure. Per a Federal Aviation Administration (FAA) regulation (FAR 25.613), if the failure consequence is critical (single point failure) to safety, then the reliability must be 99 percent with a 95 percent confidence level. For redundant components (where failure does not cause a system failure) the requirement is 90 percent reliability with a 95 percent confidence level. The requirements are based on the reliability percentage of material property distribution (failure measure) being greater than the design value with the 95% confidence level. This drives their certification work into regimes with test sampling plans in order to statistically prove reliability at a confidence level. These reliability requirements utilize a reliability index calculation that is quite similar to the QMU CF. FAA regulations (FAR 25.303) also require a 1.5 safety factor on limit loads so multiple approaches are used in complex system designs.

The motivations for devising design reliability criteria and methods that can predict failure probabilities and confidence are 1) to avoid high consequence failures, 2) to overcome design complexities and 3) to facilitate design optimizations that satisfy important constraints such as reliability, economy and small sample size. Large civil structures (bridges, buildings, etc.) are a good example of the first motivation (high consequence to failure). Very complex electronics systems (with 10K plus component counts) are the best example of the second motivation. In complex systems, the failure of a single inexpensive part (electrical resistor) can cause an expensive system wide failure. The Weapon Electrical Systems (WES) of nuclear weapons falls into this category. In these designs, components have specified reliability requirements in order to meet a requirement for overall system reliability. The nuclear weapon system reliability reported to the Department of Defense (DoD) every year is based on statistical data gathered during original production and annual stockpile surveillance activities.

The third motivation, which is to optimize constrained designs, deals with the issues that arise with nuclear weapon certification. The significant constraints facing nuclear weapons design is both the small quantity of test data and a requirement for high confidence level in the reliability. Within the discipline of statistics, reliability is reported with an associated confidence level. For the non-nuclear WES, there is sufficient test data to report reliability at a specified confidence level. There is simply insufficient test data to report a measured reliability at a specified confidence level for the Nuclear Explosive Package (NEP). This shortcoming in measured reliability existed before the ban on UGT. As a practical matter for reporting system reliability to the DoD, the reliability of the NEP is asserted to be 1.0 (sometimes called O-N-E).

In short summary, NEP QMU confidence factors are reported yearly by the nuclear Design Agencies (DAs) Los Alamos National Laboratory (LANL) and LLNL. Sandia National Laboratory (SNL) regularly reports weapon reliability numbers (without confidence level) in Weapon Reliability Report. At the beginning of nuclear stockpile assessments both reliability and a statistical confidence level were reported annually to the DoD [18]. That changed in the mid-sixties, and confidence level reporting was dropped. An excellent quote from the Love report [19] states: “No sound statistical approach exists for measuring uncertainties associated with the subjective combining of data; confidence limits could only be a measure of sampling variability, and it was considered desirable to avoid implying that uncertainties in judgments had also been measured.” Since then, SNL only reports weapon system reliability.

The SNL reliability evaluation is based on reliability logic block diagrams with reliabilities for subsystems and components [6, 20]. The LLNL Engineering Design Safety Standards, Section 11.7 [21] provides component ratings with a list of internet references. The nuclear DAs supply reliability without confidence level assessments to SNL for some select components, such as detonators. SNL incorporates the limited nuclear DA reliability data into the warhead reliability. The use of confidence level statistics is

used in design processes for components that have reliability allocation requirements. When these components are designed with sufficiently large design margins, and with planned sample size testing, there is a measure of assurance (confidence) that the design will meet reliability requirements. Typically, a majority of the component reliability testing is done during development and the full sample size of testing is completed during surveillance of the active stockpile.

While LLNL provides the system level QMU confidence factors for the NEP, the QMU methodology is being expanded and applied to NEP subsystems and components using the functional requirements of a Systems Engineering approach.

QMU Basics

The formulas and illustration of QMU are shown in Equations 1 - 4 and Figure 1. The ratio of Margin to Uncertainty defines the Confidence Factor (CF). The blue curve is a normal Gaussian probability density distribution curve to represent a limiting threshold function response. The red curve is a normal Gaussian probability distribution curve to represent a nominal function response. Each distribution has a mean μ, and a standard deviation σ. The QMU formula (4) is based on two independent distributions. It is customary to define uncertainties U as equal to one standard deviation σ of the distribution. The yellow shaded area below the intersecting distribution tails equals the probability of failure. The use of Gaussian distribution curves is for convenience. While the distributions can be arbitrary, it is most important to define a large performance margin so that that there is a minimal region of probable failure.

$$Confidence\ Factor\ (CF) = \frac{Margin}{Uncertainty} \tag{1}$$

$$Margin = M = \mu_N - \mu_T \tag{2}$$

$$Uncertainty = U = \sqrt{\sigma_N^2 + \sigma_T^2} \tag{3}$$

$$CF = \frac{(\mu_N - \mu_T)}{\sqrt{\sigma_N^2 + \sigma_T^2}} \tag{4}$$

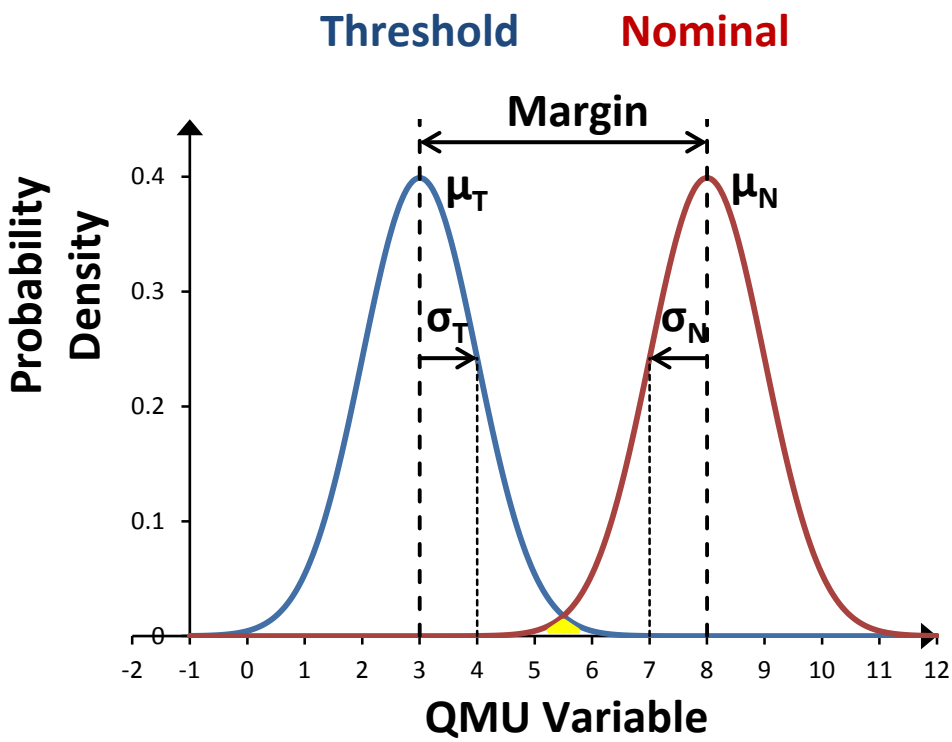

Figure 1: QMU performance and failure distribution Curve

The CF for the example shown above is 3.54 (5/1.41). The performance margin is 5 (8-3) and the uncertainty is 1.41 (square root of 1+1). Confidence factors greater than 3 are in the regime of high confidence and less than 2 are lower confidence. Since QMU is established on the basis of a single one-dimensional failure mode, its simplicity presents some key limitations. Specifically, multiple failure modes, interactions between failure modes, numerous uncertainty sources, constraints and other distribution nonlinearities all require modifications to QMU or alternative methods. However, the simplicity of QMU is its advantage. The first priority should be given to quantifying the margin. Second should be to estimate or bound the uncertainty. It is more difficult to characterize uncertainty than it is to characterize margin, so conservative uncertainty estimates are prudent. The third step depends on the CF magnitude. For CF less than 2, additional detailed M and U assessments should be pursued since there is greater potential for CF changes that could drop below 1.

Margin

The performance margin is the difference between the mean of the nominal response and the mean threshold response. As the word threshold implies, if the nominal response does not satisfy the threshold response then the performance margin is negative and the function fails. Sometimes, the performance margin is also (pessimistically) called the failure margin. Depending on the QMU application to a particular failure mode, the nominal and threshold distributions may be larger or smaller than the other. Material failure margins occur when a nominal operating condition exceeds a threshold value. In this case the threshold distribution is to the right (greater than) of the nominal distribution. Amplification failure margins occur if the nominal input (energy, temperature, voltage, etc.) triggers do not exceed threshold ignition levels. Activation energies and small sparks can trigger the

release of enormous energy outputs. The margin shown in Figure 1 is for the case where the nominal distribution is greater than the threshold. Whichever the case, the performance margin is meant to probabilistically capture a failure mode.

QMU has two key rules to govern the Margin quantification. First, a Margin is measured between the means of the threshold and nominal distributions. Second, the two distributions should be taken at the "worst" Stockpile-to-Target-Sequence (STS) conditions, i.e., the conditions associated with the smallest CF. For example, some failure modes may vary with temperature, with cold being the worst condition. In this situation, the CF is assessed at the coldest temperature. Likely environmental probabilities are not factored into nominal performance distributions. Also, to assess the smallest CF means that all environmental combinations are considered.

QMU Variable

The domain of QMU analysis is the QMU variable. After identifying a failure mode, selecting a QMU variable to characterize the failure mode is the next most important decision. The chief selection criterion for a QMU variable is that it should directly affect both the performance and the intrinsic cause of failure. Consider the example of a part that breaks under mechanical loading and causes a functional failure. The QMU variable might be the loading force or the part stress. A study of mechanical failure problems show that applied part loads cause parts to stress and that leads to part deformations or fracture failures. Other factors affecting failure includes part defects, variable part geometries and material properties that affect both the material stress levels and the material failure strength levels. More importantly, failure strength properties are the most probable failure factor and they are intrinsic material properties that can be measured. Strength is measured in the units of stress not load. Thus in this instance, since strength[1] is an intrinsic material property and a cause of failure, it is the better choice for a QMU variable.

In the previous example, the selection of stress as the QMU variable represents the better choice over applied load. There is no single correct choice. Selecting applied load as the QMU variable is also possible and that can sometimes lead to different QMU results. Differences in results are traceable to nonlinearities in transfer-functions that relate material strengths to applied load forces (the selected QMU variable). Other nonlinearities in the uncertainty quantification also change the QMU result. QMU variables fall in three general categories. There are inputs (applied loads and displacements), there are intrinsic state variables (stresses) that cause parts and materials to behave in certain ways, and there are outputs (transferred loads, geometric motions, gaps or deflections). The QMU variable can be any one of these choices, but the best choice has the closest alignment between the intrinsic failure cause and the functional requirement.

Another key consideration when selecting the QMU variable is if it is measurable. With QMU variable measurements, it may be possible to validate the nominal or threshold behaviors and maybe even the CF itself. With the single QMU variable choice, two distribution functions need to be defined. The ability and convenience to make physical measurements of either nominal or threshold behaviors is the best and most direct path to doing QMU. The next best consideration is if the QMU variable is predictable, either analytically or computationally. Taking a predictive approach introduces additional uncertainties. Finally, the next consideration is that independent uncertainty sources should be affecting the variance of nominal or threshold performances. Stated differently, if there is a significant uncertainty source that affects the performance margin, then that uncertainty should be quantifiable using the QMU variable.

[1] Stress works well for quasi-static ductile failure, but brittle material failures or long term creep damage may have other better choices for the QMU variable

Uncertainty

Performance uncertainty is assessed as the Root-Sum-Squared (RSS) of independent nominal and threshold uncertainties. The nominal U_N and threshold U_T uncertainties are each composed of independent parameters that are also calculated using a RSS (Equations 5 and 6). N_N and N_T are the number of independent uncertainty contributions from nominal and threshold distributions.

$$\sigma_N = U_N = \sqrt{\sum_{i=1}^{N_N} U_{N_i}^2} \tag{5}$$

$$\sigma_T = U_T = \sqrt{\sum_{i=1}^{N_T} U_{T_i}^2} \tag{6}$$

In a QMU assessment, the uncertainty sources are identified and then quantified. The following list provides a general guide to aid identification:

1. Material properties
2. Fabrication, manufacturing, assembly and life-cycle use
3. Experiments and testing: repeatability and reproducibility
4. Stochastic physical phenomena
5. Analytical and computational
6. Unknown-Unknowns

Before going into more detailed descriptions regarding the variety of uncertainty sources mentioned above, it is important to note that in the many applications of QMU to date, the nominal and threshold distributions are constructed using normal Gaussian (μ, σ) probabilities. From experiments and simulations, the nominal and threshold σs are measured or predicted. For the extraneous sources of uncertainty (#5, 6), their contributions are added to conservatively bound the total uncertainty.

Uncertainties in material properties are typically in the range of 5-10%. Simple statistical formulas for mean and standard deviation are used for characterizations. Assume there is a set of material property measurements x_i which define a normal distribution X. The mean $\bar{X}$ is given by Equation 7 and the standard deviation s of the data is Equation 8. The testing sample size is n. The coefficient of variation η is defined by Equation 9 and equals the standard deviation (uncertainty) divided by the mean. The η for the strengths of common structural steels ranges between 0.05 and 0.07 [7].

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7}$$

$$s = \left[\frac{1}{n-1}\sum_{i=1}^{n}(\bar{X} - x_i)^2\right]^{\frac{1}{2}} \tag{8}$$

$$\eta = \frac{s}{\bar{X}} \tag{9}$$

If the QMU variable is stress and the failure mode is stress exceeding strength, then the uncertainty in the yield strength mentioned above would be the standard deviation calculated in Equation 8. Engineers are accustomed to taking material property values from a variety of sources and then using them conservatively in their work. In some industries, material measurements are taken from material build

lots used in manufacturing so the data for actual material uncertainties is available. When actual property data is unavailable, the variation in coefficients of similar materials can be used to estimate the material property uncertainties.

In the category of fabrication uncertainties, manufacturing variations influence both the threshold and nominal distributions. Examples of the numerous factors that affect fabrication uncertainties are the selection of materials, suppliers, their manufacturing process history, humidity, baking and degassing, cleaning solutions and even to how parts are stored during different fabrication stages and then on to in-use storage conditions. Parts stored on a shelf frequently differ in their behaviors from parts that endure a service lifetime with customers. The variety of starting materials, properties and environments can physically alter the material behavior (stress-displacement) of explosives, ceramics, polymers, etc. It is through inspection that fabrication uncertainties are reduced. Whether by measuring part dimensions or by destructive sample testing, it is highly desirable to only use parts that possess acceptable (superior) characteristics. Eliminating (by truncating) the undesirable tails in a performance or threshold distribution through acceptance inspection favorably affects the resulting failure probability distribution. By truncation or elimination of first-order characteristics that contribute to failure, the CF is increased along with part reliability. Second-order contributions to uncertainty remain (inspection errors, gaging errors, part distortions, material variations within a lot, etc.)

Fabrication is typically the largest uncertainty source, and it can be broken down into two types: aleatory and epistemic. Aleatory uncertainty, meaning stochastic randomness, happens in most manufacturing processes. It is generally believed that establishing tight process controls will minimize manufacturing variations, and component quality screening can limit as-built variations further. Physically random processes will introduce some unavoidable aleatory uncertainty.

Epistemic uncertainty, meaning imperfect knowledge, happens because parts and processes are imperfectly characterized. A significant source of epistemic uncertainty (undetected) occurs in assemblies where there is mechanical interference (or gaps) between parts, such as slide fittings where shafts fit into part holes. The dimensions of the holes and shafts can each have dimensional tolerances that are characterized by Gaussian distributions with means and standard deviations. There will be some combinations of holes and shafts where the part dimensions interfere with the similar overlap region as is studied in QMU. If these slide fittings are contributing to part failures, then they are potential sources of uncertainty. Note that variations in gaps structures are just as important as potential uncertainty sources especially when there is lubrication involved and when the lubrication changes over time.

There is a rich scientific body of work describing experimental and test measurement uncertainties [8, 22]. The simplest rule of thumb is to ensure a high ratio between the measurement and the measurement uncertainty [23]. When making test measurements, there are uncertainties with the measurement itself, the test setup, the repeatability and reproducibility of test measurements and many other factors which should be assessed (human operators, environmental factors, data processing, quality, etc.). There are numerous analysis of variance method models that determine and ascribe uncertainty to the experiment factors. A two factor study with replications called Gauge Repeatability and Reproducibility (Gauge R&R) looks at measurements made by different operators on multiple parts. It is important to understand the uncertainty associated with the parts as opposed to uncertainty in the operators.

Stochastic physical phenomena are those processes whose outcomes are seemingly random events. Very often these processes are modeled using Monte Carlo simulations: radiation transport is one such phenomenon. Sometimes when a subsystem is treated as a black box with no attempt to scientifically

characterize the physical phenomena and hence deduce the behavior into a deterministic outcome, then that subsystem has measurable random outputs that can be statistically characterized.

The uncertainties attributed to analytical and computational sources [24, 25] encapsulate a multitude of uncertainty sources. Since modeling parameters include material properties, fabrication, testing configurations, and include stochastic phenomena, the four previously discussed uncertainties are all potential factors that can be included in modeling. The particular uncertainty that is unique to analysis and computations is the accuracy of that modeling. Perhaps the two largest uncertainty pieces are the modeling idealization (the discretization and modeling simplifications: a spherical chicken with uniform heat flux) and the simulation physics (are the right equations being solved correctly?).

There is a testing uncertainty analog to this analytic and computational uncertainty. Test unit configurations may not exactly match weapon configurations so there is an error in that representation that introduces uncertainty to the test results. Environmental ground test conditions of temperature, vibration and shock may not excite or induce the same physical response of a flight test unit experience so there is uncertainty in the accuracy of the physical response.

With capable analysis and computational tools, there is an opportunity for uncertainty quantification. Sensitivity analysis can guide the uncertainty quantification efforts to focus on the most important modeling assumptions and inputs. Modeling can examine how variation in a particular uncertainty source affects the output of a QMU variable. Through brute force, deterministic simulation capabilities can be adapted to take probabilistic inputs and predict probabilistic outputs. Thus with probabilistic analyses, it is possible to estimate nominal and threshold (μ, σ) response surfaces. As a general caution, probabilistic analyses are ill suited to predicting distribution shapes and the low probability magnitudes of distribution tail behaviors.

When quantifying experimental and computational uncertainties, additional uncertainty contributions are added to the QMU calculation. The consequence of additional uncertainty contributions to QMU is shown in Figure 2. Additional uncertainty contributions increase the failure probability by increasing the overlap in the failure zone (from yellow to red). In QMU, there is no presumption that the shape of the distribution curves must be known. Instead, a Gaussian is assumed for convenience and the goal is to identify and then to conservatively estimate an uncertainty bound.

The inclusion of numerous uncertainty sources in U for CF is what differentiates QMU from the analytic, reliability index methods. In those well-defined problems, the uncertainty contributions are limited to known problem variables being quantified. With these restrictions, the distribution shapes can be mathematically transformed and combined to determine a well characterized probability of failure distribution. QMU prioritizes a conservative CF estimate that includes all potential uncertainty sources over distribution shape predictions.

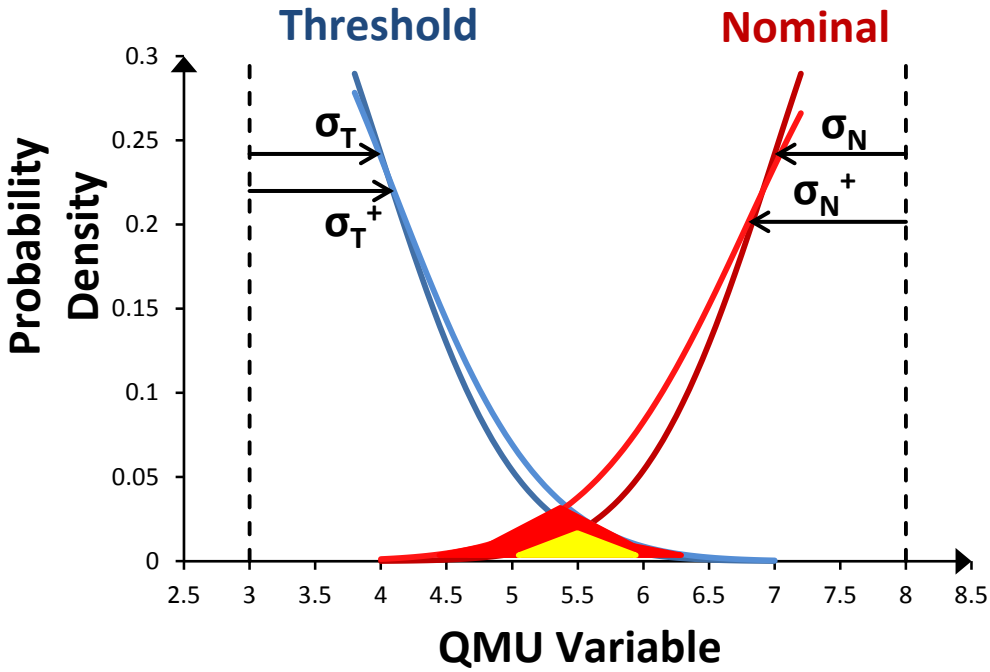

Figure 2: With a 10% increase in Threshold uncertainty and a 20% increase in Nominal uncertainty, the failure probability increases (from yellow to red area) and the QMU CF decreases by 13%.

The acknowledgement of "unknown-unknowns" is an important topic in QMU because when over-looked, it can lead to over confidence in a design. Studying the application of QMU reveals three unknown-unknown situations: estimation errors in uncertainty or margin, not recognizing a source of uncertainty, and not recognizing a potential failure mode. Analytic, experimental and computational improvements in uncertainty quantification can minimize error uncertainties. Ever improved scientific and engineering studies can help to minimize a failure to recognize uncertainties and potential failure modes. While it is possible to bound uncertainty estimates, the failure to recognize potential failure modes is the most serious of the unknown-unknowns. Since a QMU analysis is predicated on the study of a failure mode, an unknown failure mode implies an incomplete or worse incorrect QMU problem is being studied.

It was stated previously that the shape of the threshold and nominal distribution curves is chosen to be a normal Gaussian for convenience. It is also customary to use one standard deviation when reporting a QMU CF. Thinking rhetorically, does the QMU practitioner know the distributions well enough to accurately represent the shape of the output distributions down into the regime where the tails of threshold and performance distributions overlap? Secondly, at what standard deviation factor does the QMU practitioner have the better uncertainty estimate, one, two or three standard deviations? Because tail characteristic behavior is in the statistical regime of low probability events, and because most often there is not enough data or analysis certainty to definitively quantify tail probabilities with high confidence, there can be misplaced emphasis and energy put into obtaining accurate tail behavior predictions.

The nuclear weapon QMU experience suggests that uncertainties can dominate the CF ratio and which can lead to a significantly lower CF calculation. Over time, uncertainties tend to be reduced and the CF increases. The experience with legacy weapons is that improvements in CF are mostly the result of better uncertainty characterization. While, for refurbishment Life Extension Programs (LEPs) and Alterations (Alts), improved CFs are the result of purposeful, higher design margin improvements. The conditions of lowest CFs typically occur where the margin also happens to be the smallest. The variation of uncertainty magnitudes over the range of possible STS conditions is a second-order factor when compared to the change in margin over these environments. Instead of uncertainty variations as the second largest source of low CFs, the age of a weapon or its subsystems can be more important. So "Old and cold" is a common phrase that captures the two common worst case conditions under which QMU is assessed.

Systems Engineering Approach to Nuclear Weapon Stewardship

LLNL tailors a Systems Engineering approach to nuclear weapon stewardship that is drawn from DoD [26] and NASA [27] experience. The method is based on a thorough definition of functional requirements that starts with Customer requirements, translates them into System level functional requirements and then breaks these down into derived functional requirements that support higher level functions. The benefits of this approach are its comprehensiveness across all weapon functions and its consistent application to all weapon functions. The flow of requirements from the Customer to the weapon system to the derived requirements is called traceability. The reason this paper describes nuclear weapons, systems engineering and requirements is because the application of QMU to weapon assessments is made on this basis. The systems engineering approach defines functional requirement details and relationships.

Figure 3 is a graphical representation of a nuclear weapon functional requirement breakdown. For obvious classification reasons, key functional details and the functional relationships are omitted. An essential characteristic of the functional requirements is their hierarchical flow. The DoD defines the Customer requirements (top oval of Figure 3). The three main sources of customer requirements are Military Characteristics (MC) documents, Interface Control Documents (ICD) and environmental conditions documented in a Stockpile to Target Sequence (STS) report. The MC defines the major warhead requirements and the ICD defines geometric constraints, power, electrical signals, handling equipment and transport gear. Customer Requirements are negotiated between the DoD, NNSA and the weapon design laboratories.

The nuclear Design Agency (DA) LLNL and the non-nuclear DA SNL define System requirements from Customer Requirements. Below the System requirements are the derived requirements. These flow down from System requirements. Each of the weapon functional requirements is captured in a rectangular box. The base expectation of a functional hierarchy is that when lower level functions are satisfied, then the higher level function is also satisfied. In the example below, when the mechanical interfaces, and weapon electrical system functions are satisfied, then the warhead is integrated within the DoD system.

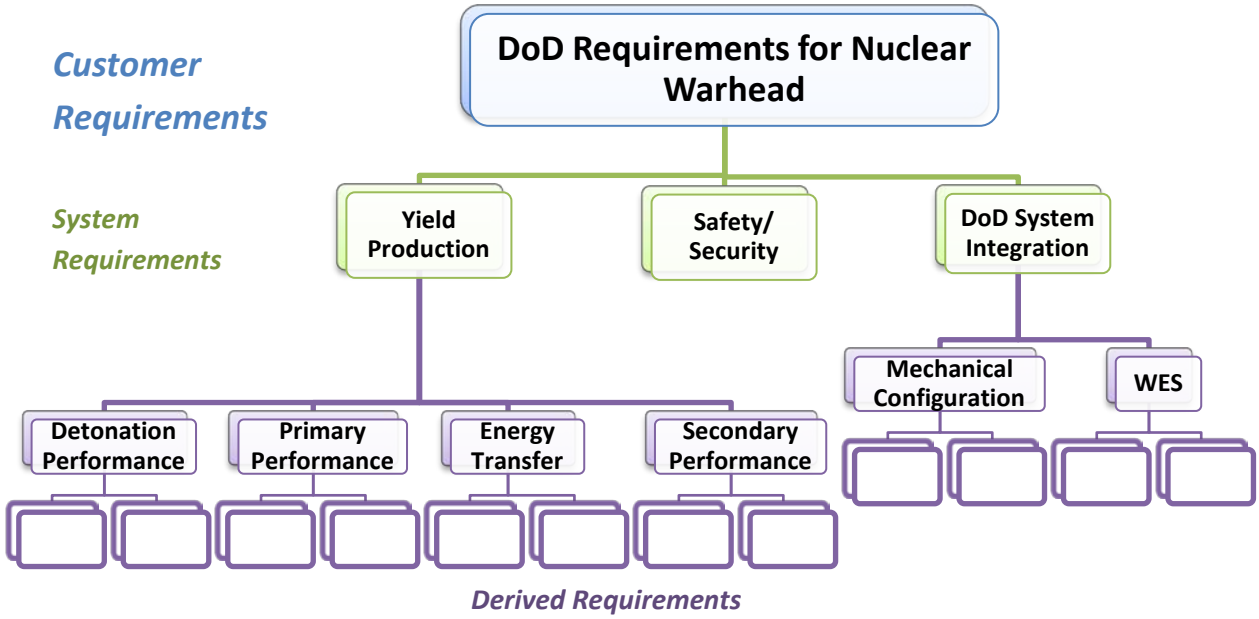

Figure 3: Hierarchical Functional Requirements

Each of the boxes shown in Figure 3 shows a functional requirement label. Each function box has an associated requirement statement that is carefully worded to define "how well something performs" (see Figure 4). The "something" in a functional requirement represents the design solution. Just as there is a functional requirement hierarchy, there is also a physical design hierarchy. The physical system hierarchy starts with the full system at the top drawing or "design definition", and then defines every subsystem and component below. The design solution is intended to meet the requirements specified in the functional requirements hierarchy.

Well-defined requirements [27, 28] are very important to Systems Engineering: both during the design development period and in the sustained stockpile stewardship period. Requirements need to be unambiguous and measurable. The following is an example of an unambiguous and measurable requirement: the nuclear weapon shall produce X Kilotons of nuclear energy output. The "something" is the nuclear weapon. The "what" is nuclear energy output and the performance requirement is measured in X Kilotons.

A typical nuclear warhead has about 100 derived functional requirements. For the design solution to each requirement function, there is a list of Performance Parameters (PP) that can affect how well the function is performed. PPs include any characteristics (material properties, physical dimensions, and state characteristics) that affect the functional performance. Across all functions, the total quantity of PPs can be upwards of 500. For each function, a subset of the total number of PPs is judged critical to performance. Critical Performance Parameters (CPPs) are associated with potential failure modes. By careful examination of CPPs, the number of potential failure modes per function is reduced. The cause and effect expectation between CPPs and failure modes is that if CPP specifications are not met, then the functional performance is critically jeopardized. Since the PPs affect functional performance, they are key sources of uncertainty. PP and CPP variations affect the immediate function, as well as all the functions above them.

Most CPPs vary within a specified range because they were (or should have been) specified at the time of original production. Even when specified, fielded weapon designs have “as-built” characteristics that can be studied to provide better variance information. The variances of CPPs are typically the major uncertainty sources necessary for QMU analysis.

Figure 4 is an example of the information associated with a derived function. Figure 3 shows a function requirement labeled Primary Performance and it has additional unspecified derived functions below it. Figure 4 expands the Figure 3 Primary Performance function to show additional derived requirements flow-down. It contains the functional requirement, lists PPs and CPPs and then it makes a triage statement that dispositions the QMU analysis method and identifies the QMU variable. The QMU assessment process triages functional requirements into one of three assessment modes. In the three derived sub-function requirement examples of Figure 4, the QMU triage results are QMU assessment, QMU roll-up and null. The next few paragraphs will describe these further.

For a few functional requirements, there is no significant impact to failure, or there is no conceivable failure mode, or the failure likelihood is extremely improbable. For these situations, the QMU triage result is deemed a null consequence outcome, and it is decided to not perform QMU. This does not happen often, but when it does, performing QMU serves no meaningful purpose. The derived functional requirement Primary Centering Configuration is an example of a null result from QMU triage. The weapon design requires that cushions will support and center the Primary. This is considered “good engineering practice.” A QMU assessment of this function examines how the design solution (rubber cushion) might fail, what is the likelihood of failure and what would be the consequences. First it is assumed that there is no primary assembly birth defect where cushion is deformed, missing or severely damaged. Testing showed that damaging the material is extremely unlikely, either mechanically or chemically. When extreme variations in cushion properties are assumed, engineering and physics assessments show that the failure consequences at the next higher function level are acceptable. In short, unless the cushions melt or sublimate out of existence, likelihood is extremely small and the consequences are minimal. Hence the QMU triage result is null. In these situations, the prudent course is to monitor the function for unknown and unexpected failure modes.

The Defects function example involves a chemical mitigation problem. A magic material controls chemical reaction degradation rates in the Primary and slows the rate of defect growth. An acceptable defect mass is established and assessed as an uncertainty in the higher Primary Performance function. At the derived Defect function level, a QMU assessment is performed to assure that defect growth is acceptably small over a weapon lifetime. Since a maximum defect mass requirement is established, it is decided to re-pose the problem in the terms of the defect growth rate over a lifecycle period of time. The selected QMU variable is years. The threshold distribution is based on how many years it takes for a defect to grow to the maximum allowable mass. The nominal distribution is weapon age, from the dates of manufacture to weapon removal from the stockpile.

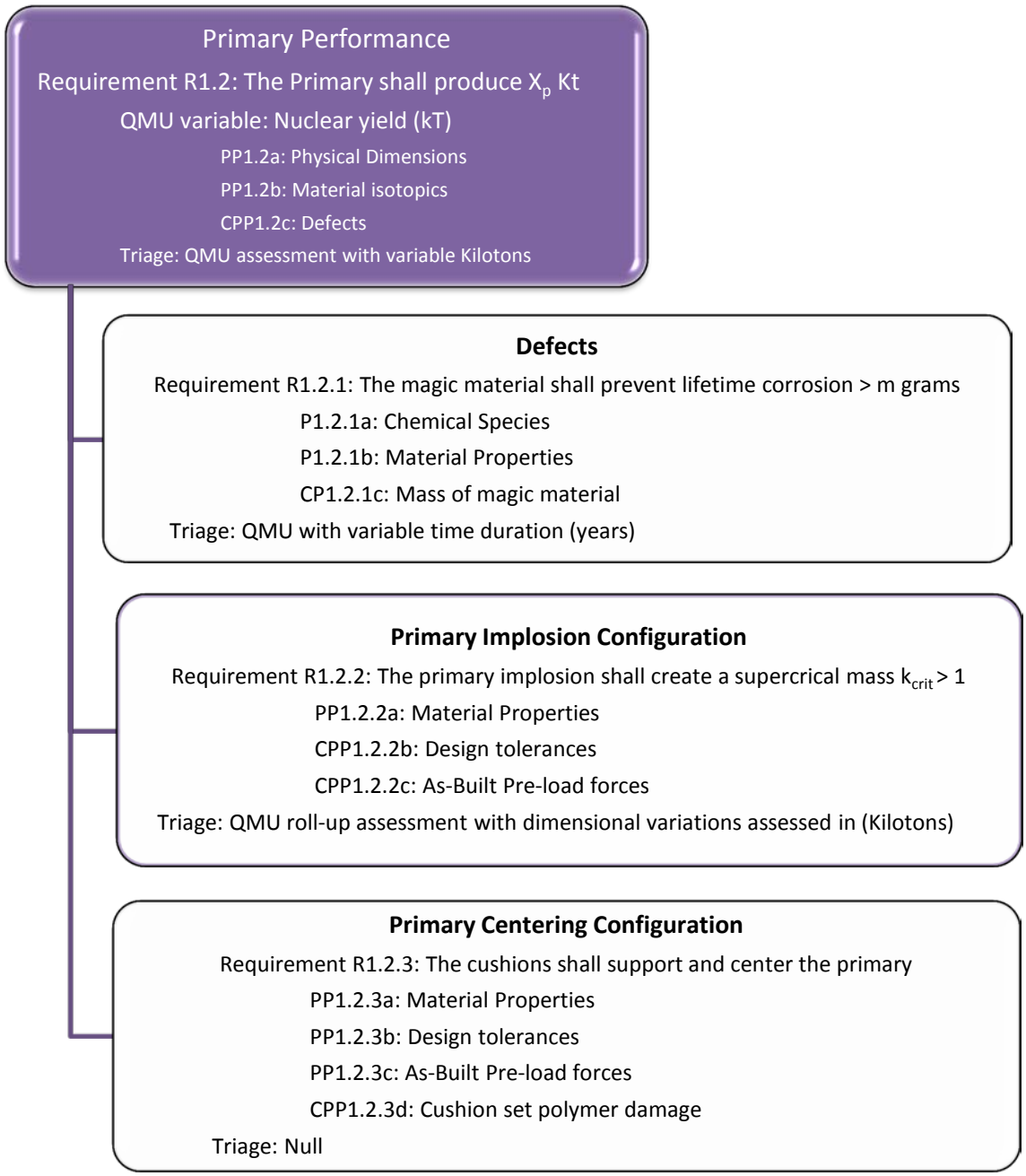

Figure 4: Example of derived functional requirements and QMU variables with associated performance parameters

The Primary Implosion Configuration derived function example involves two common issues for engineering QMU: physical configurations and structural integrity. Structural integrity is the assurance that nothing breaks, deforms or moves in an unintended way. Physical configurations are an external measure of structural integrity. For both structural integrity and physical configurations, the failure risk is in regards to likelihood and consequence. For some components, fracture and cracking would be a cause for system level failure. For other parts (rubbery cushions for instance), cracks do not lead to a system failure. There is a similar situation of ambiguity for plastic deformation, adhesive de-bonding,

migration of chemical species, gap formation, distortions, etc. Physical configurations and structural integrity are clearly important PPs, but it is not always the case that they are CPPs.

Physical deformations, part movements, gap formations and shifting components are assessed by engineering for structural integrity, but they are also routinely assessed by physics to determine impacts to higher nuclear requirement functions. Sometimes, physics sets physical configuration limits that can be used by engineers as threshold design requirements; but more frequently, physical configuration limits are intended to serve a notification purpose. When weapons deviate from design definitions, engineering and physics assess the structural integrity and nuclear performance respectively. In the circumstance where a lower, derived functional requirement has low failure risk, but the higher functional requirement is affected then a QMU roll-up assessment occurs. In the Primary Implosion Configuration requirement example, the risk of failing the criticality requirement is judged inconsequential, but the dimensional configurations are judged important to the higher Primary performance function. Hence the triage result is QMU roll-up assessment with a QMU variable of Kilotons.

The novice to QMU might have expected the QMU roll-up variable would be a length measure such as mm. Instead the variable is Kilotons because that is the chosen QMU variable at the higher functional requirement level. So the QMU analysis must characterize dimensional variations and assess them as Kiloton variation uncertainties.

As a last observation, there can be multiple failure modes per functional requirement. Using the Primary Configuration requirement example of Figure 4, as-built pre-loads are identified as a performance parameter. This is referring to assembly loads that can stress parts, cause gaps between parts and otherwise affect Primary Performance. If there is an intermediate to high risk failure mode associated with loading, then QMU analysis would be warranted. Alternatively, the variation in as-built loads could be rolled up and analyzed as an uncertainty to Primary Performance. With multiple QMU analyses and multiple roll-up of performance parameter variances, there are many permutations of possible approaches. One recommendation about multiple QMU outcomes is to re-examine the quality of the functional requirement definition and ensure that the requirement is unique and measurable.

QMU Assessment Process

The QMU Assessment process itself is broken into two parts, one for QMU failure mode assessments and the second for QMU analysis. Before defining, these two processes, it is important to understand how QMU fits in the nuclear weapon stewardship context at LLNL. Figure 5 shows that QMU failure mode assessments are one portion of the Performance Risk Management process and which is itself one of the six Annual Assessment Review processes: surveillance analyses, design departure analyses, performance risk management and QMU, aging basis management, reliability analysis and knowledge gap management. These processes are executed continuously and the results reviewed yearly. QMU failure mode assessments are also triggered when there is a Significant Finding Investigation (SFI). Most SFIs occur from surveillance Disassembly and Inspections (D&I), component testing and Joint Test Assembly (JTA) flight testing. There have also been SFIs initiated from computational assessments.

As part of Annual Assessment Review (AAR) processes, performance risks from numerous sources are evaluated. The inputs to this review process are many, surveillance observations, component evaluations, subsystem testing, chemical and radiological experiments and analyses from many disciplines. A QMU analysis is decided based on the high to intermediate priority performance risks.

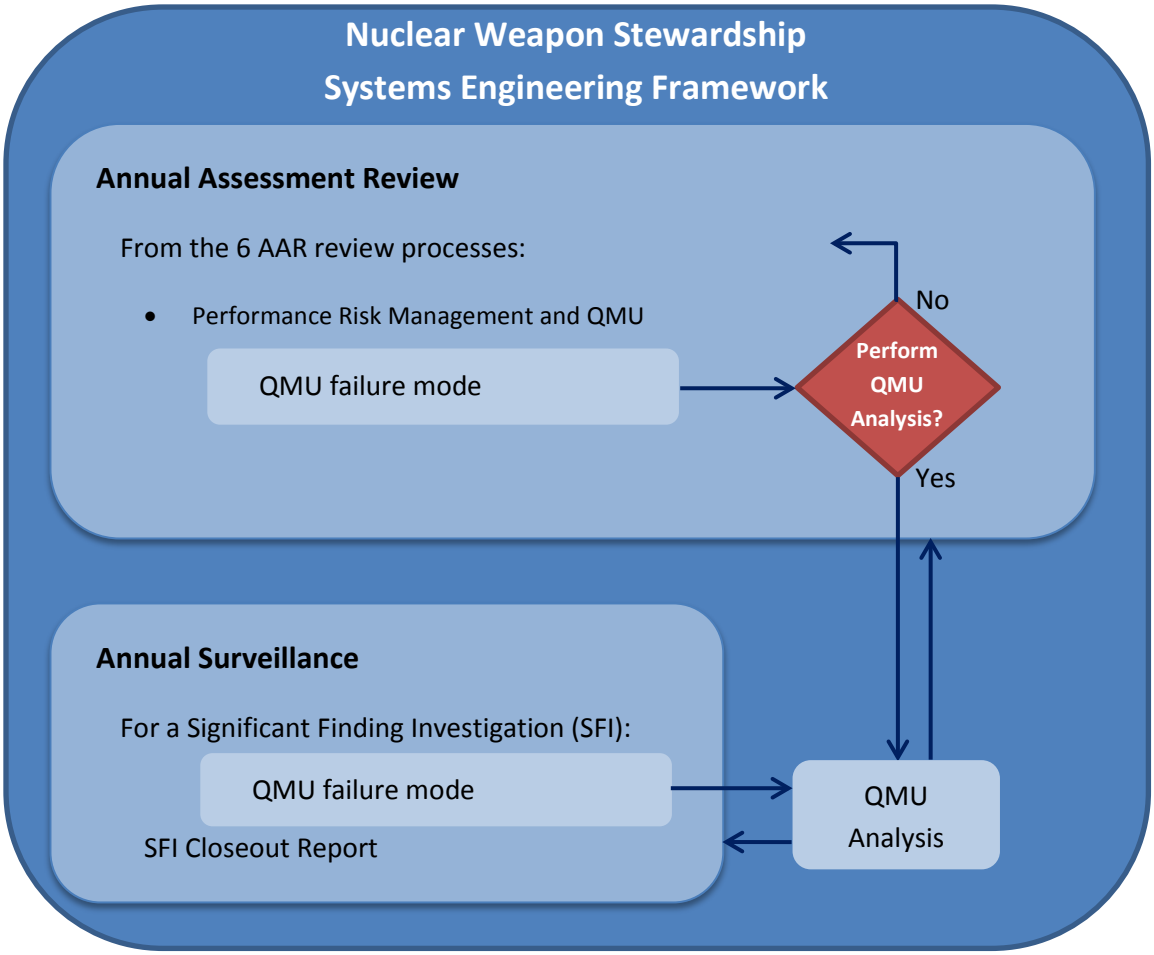

Figure 5: QMU Assessment Processes that occur within Nuclear Weapon Stockpile stewardship

A prerequisite of QMU assessments is the preparation of the nuclear weapon functional requirements hierarchy. These functions identify the potential "so what" consequence when failures occur. As risk captures the consequence and likelihood of functional failures, it is the best process by which to trigger QMU analyses. Performance risk assessments examine the collected set of possible failure modes, decide on the most likely failure modes based on available information, and then consider the consequence of function output variation and how the functional output impacts higher level functions.

For SFIs, the focus is on studying anomalous conditions, observed or simulated. Sometimes, the anomaly is completely unexpected, or it is related to a PP or CPP that is out of specification. Sometimes the anomaly is a failure (possibly realized risk) in a functional requirement output. The QMU methodology is a rigorous approach for studying how anomalies affect functional requirements. QMU analysis is expected as part of the SFI closeout report.

The two parts of QMU assessments are described in the following process steps.

QMU Failure Mode Assessments

1. Review the functional requirement, identify performance parameters and determine the CPP subset for further study.

2. Study all potential failure modes to the function. Consult SMEs, review history, compare to similar systems, brainstorm and prioritize failure modes based on risk with likelihood and consequence.
3. Reexamine all PPs and CPPs known to affect the function and its associated failure modes and update the body of system engineering information for the function.
4. Triage the failure modes of the functional requirement. Using expert judgment, considering failure risks, design, test and experimental margins, judge the validity of the failure mode and the benefit of further QMU assessment. Consider how performance parameters affect the failure mode. Decide if the failure mode is suitable for QMU analysis, should be rolled up as uncertainties to QMU in a higher level functional requirement, or decide if the QMU outcome is null.

QMU Analysis

1. Start QMU analysis with examination of possible QMU variables for the failure mode. At the beginning, it is helpful to make multiple choices and down-select after further assessment information is available.
2. For each of the QMU variable choices, quantify their margins with available data and use of expert judgment as necessary. Identify all possible sources of uncertainty.
3. Identify gaps in knowledge and lay out a functional assessment plan that will quantify the performance and failure data necessary to compute QMU. This plan should strategize the methodology for determining threshold and performance distributions along with the statistical quantification techniques.
4. Perform the QMU calculations to determine CF. Most QMU assessments are not performed in a linear step-to-step process, but instead require looping back to reassess failure mode risks, reconsideration of QMU analysis versus roll-ups of parameter variances to higher level functional requirement QMUs. There are multiple options for QMU variable choices and the methods for CF quantification may have to be altered based on assessment simulation and test results.

A performance risk assessment process examines the collected set of possible failure modes, decides on the most likely failure modes based on available information, and considers the consequence of function output variation and how the functional output impacts higher level functions.

Conclusions

This paper has reviewed the basics of QMU analysis, the margin, QMU variable and uncertainty. QMU is rooted in the simple design principle that performance margins should exceed design uncertainties. The quantification of uncertainty should consider all potential sources both known and unknown. The application of QMU is based on a system engineering approach applied to functional requirements and their numerous potential failure modes.

Based on QMU experiences, the following are basic QMU practice rules:

1. Performance margin is defined between the means of the threshold and nominal distributions
2. Unless documented otherwise, uncertainty is based on the RSS of one standard deviations
3. The QMU variable should be selected from performance parameters associated with the threshold failure mode and be measurable, intrinsic properties.
4. The CF is calculated at the worst possible “normal” conditions as specified in the STS. The CF does not take credit for the likelihood of environmental or logistical conditions.
5. Apply conservatism to the quantification of uncertainties

The QMU CF metric measures the ratio of a design performance margin between nominal and threshold distributions over uncertainties. While there are similarities between the probabilistic reliability index and the QMU confidence factor, they are not the same. QMU applies a more conservative assessment because it considers a larger and wider range of potential uncertainty sources. Additionally, QMU does not assume that the distributions shapes of nominal and threshold are known. QMU assumes Gaussian normal distributions for convenience, one sigma uncertainty estimates, adds additional uncertainty terms, and considers worst possible conditions for the assessment so that meaningful and conservative design confidence statements can be reported.

Appendix A

Statistics and Reliability Design Methods

Throughout this paper, there has been a purposeful avoidance of statistics and other reliability design methods. While these methods can be technically challenging and powerful assessment tools, they distract the novice QMU practitioner from gaining the basic understanding of how QMU is applied in nuclear weapon stockpile stewardship. The more analytic and technical approaches to assessing reliability and confidence have been deferred to this Appendix. The statistical treatments presented here are rudimentary and without derivation. Tables and graphs are presented of simple statistical concepts with a purpose to show the mathematical relationship between reliability and confidence. The purpose here is to provide a basic information review of statistical reliability, confidence levels and reliability based design methods and to encourage further training and application of statistical methods in stockpile stewardship. These statistical basics better inform QMU statements.

With Yogi Berra like wisdom, the French and German Science academies noted that "Probability-based statements are per se fraught with uncertainty" [30]. Their major point is the communication difficulties between science and the public. The subtle "per se" mathematical point is that reliability statements should be made with an associated confidence level statement and that is not always possible or advisable given an audience that does not have statistics training. There have been several proposals to create scales and standards for communicating scientific probabilities and uncertainties [31] for the purposes of communication consistency and clarity plus legal indemnification. A range of descriptive words are assigned to levels of probability with associated expectations for action. Nuclear weapon hazard analysis uses probability scales in screening tables [32]. Without common standards, communicating probabilities and statistics to an uninformed audience is counter-productive. Generally, statistical statements do not clarify risks and uncertainty for the public or most scientists.

The first of six topics in this Appendix reviews how to algebraically manipulate Gaussian distributions. With this knowledge, the topic of First Order Reliability Methods (FORM) is introduced and it shows that QMU solutions are not unique, but dependent on multivariate (QMU variables) solutions. Then in the third topic, the CF is related to reliability, confidence level intervals and sample sizes. The fourth and fifth topics cover Safety factors and the Sandia National Laboratories (SNL) K Factors. The K Factor method is a practical methodology that can be framed in a QMU context and used in product development and qualification processes. It includes specifications for design reliability at a specified confidence level with sample size guidance for testing. Lastly, measured reliability is presented to show the quantity of testing necessary to demonstrate high reliability in components. This last topic segues to Appendix B where nuclear weapon surveillance is discussed in the context of 90% reliabilities at 90% confidence levels.

Algebra of Normal Functions

The Haugen [5] reference contains a chapter entitled "Supporting Mathematics" to provide an analytic approach to calculate formulas composed of statistical functions. As in the rest of this document, only normal distributions are considered. For a formula F=A(B+C)/D, each of the four functions A - D can be described with means (μ_a, μ_b, μ_c, μ_d) and standard deviations (σ_a, σ_b, σ_c, σ_d). The problem is to derive the mean and standard deviation of F in terms of A - D means and standard deviations.

For the sum of two normal distributions, the addition results in a new distribution with a mean μ_{sum} and a standard deviation σ_{sum} (Equations A.1 and A.2). For the difference between two distributions, the individual means are subtracted from each other and the standard deviation is the same as A.2.

$$\mu_{sum} = \mu_b + \mu_c \tag{A.1}$$

$$\sigma_{sum} = \sqrt{\sigma_b^2 + \sigma_c^2} \tag{A.2}$$

When two distributions are multiplied together, the new distribution mean μ_{prod} and standard deviation σ_{prod} are given by Equations A.3 and A.4. In similar fashion the quotient of two distributions are defined in A.5 and A.6. The combination of a sum, product and division provides the final result for the formula mean μ_f and standard deviation σ_f (Equations A.7 and A.8).

$$\mu_{prod} = \mu_a \cdot \mu_{sum} \tag{A.3}$$

$$\sigma_{prod} = \sqrt{\mu_a^2\sigma_{sum}^2 + \mu_{sum}^2\sigma_a^2 + \sigma_{sum}^2\sigma_a^2} \tag{A.4}$$

$$\mu_{quot} = \mu_{prod}/\mu_d \tag{A.5}$$

$$\sigma_{quot} = \sqrt{\frac{\mu_{prod}^2\sigma_d^2+\mu_d^2\sigma_{prod}^2}{\sigma_d^4}} \tag{A.6}$$

$$\mu_f = \mu_a(\mu_b + \mu_c)/\mu_d \tag{A.7}$$

$$\sigma_f = \sqrt{\frac{(\mu_a\cdot(\mu_b+\mu_c))^2\sigma_d^2+\mu_d^2(\mu_a^2(\mu_b^2+\mu_c^2)+(\mu_b+\mu_c)\sigma_a^2+(\mu_b^2+\mu_c^2)\sigma_a^2)}{\sigma_d^4}} \tag{A.8}$$

First Order Reliability Methods

The First and Second Order Reliability Methods, FORM and SORM respectively are analytic methods to solve complex multivariate design problems. The methodology is powerful because it is a means to perform design trade optimizations. The CF formula (Equation 1) defines a limit state function *g* for the margin between two normally distributed probability density functions Threshold *T* and Nominal *N* (Equation A.9). The failure probability P_f is defined (Equation A.10) when the limit state function is less than zero. The notation $g(\cdot) \geq 0$ denotes the failure surface to a safe region of probability.

$$g(X,Y) = T(X) - N(Y) \tag{A.9}$$

$$P_f = P[g(\cdot) < 0] \tag{A.10}$$

Per the QMU method, the CF is defined by Equation 4 given knowledge of the mean and standard deviations. As the complexity (nonlinearity) of limit state functions increase, they do not necessarily possess unique Most Probable failure Points (MPP). In the following example taken from Choi [10], it is possible to show in a relatively simple beam problem that different CFs are obtained when two equivalent limit state functions are solved. Said differently, the QMU variable selection can alter the MPP CF.

A simply supported beam, length *L* is loaded at its midpoint with a load *Q*. The beam plastic section modulus is *W* and the beam yield strength is *T*. All four variables have independent and normal distributions with known means and standard deviations (Table A.1). Two different, but equivalent, limit state (QMU variable) functions define the distribution for nominal loading and threshold failure. The first limit state function is based on the beam moment (Equation A.11) and the second is based on the yield strength (Equation A.12).

$$g_1(Q, L, W, T) = WT - \frac{QL}{4} \tag{A.11}$$

$$g_2(Q, L, W, T) = T - \frac{QL}{4W} \tag{A.12}$$

	μ	σ
Q	10 kN	2 kN
L	8 m	0.1 m
W	100×10^{-6} m^3	2×10^{-5} m^3
T	600×10^3 kN/m^2	1×10^5 kN/m^2

Table A. 1 Beam property distribution data

Using the algebra of normal functions, the following solutions are obtained for g_1 and g_2 CFs. The two results are considerably different which raises doubts regarding the certainty of knowing if the minimum CF represents an MPP solution.

$$CF_{g_1} = \frac{\mu_{g_1}}{\sigma_{g_1}} = \frac{\mu_W \mu_T - \frac{1}{4}\mu_Q \mu_L}{\frac{1}{4}\sqrt{\mu_Q^2\sigma_L^2 + \mu_L^2\sigma_Q^2 + \sigma_L^2\sigma_Q^2 + 16\left(\mu_W^2\sigma_T^2 + \mu_T^2\sigma_W^2 + \sigma_T^2\sigma_W^2\right)}} \tag{A.13}$$

$$CF_{g_2} = \frac{\mu_{g_2}}{\sigma_{g_2}} = \frac{\mu_T - \frac{\mu_Q \mu_L}{4\mu_W}}{\frac{1}{4}\sqrt{16\sigma_T^2 + \sigma_{\frac{QL}{W}}^2}} \tag{A.14}$$

$$\sigma_{\frac{QL}{W}} = \sqrt{\frac{\left(\mu_Q \mu_L\right)^2 \sigma_W^2 + \mu_W^2\left(\mu_Q^2\sigma_L^2 + \mu_L^2\sigma_Q^2 + \sigma_Q^2\sigma_L^2\right)}{\sigma_W^4}} \tag{A.15}$$

$$CF_{g_1} = \frac{40.00}{16.25} = 2.46 \tag{A.16}$$

$$CF_{g_2} = \frac{400 \times 10^3}{115 \times 10^3} = 3.48 \tag{A.17}$$

Invariant solution methods to these problem types are found in the work of Hasofer and Parkinson [33, 34] outlined below. They require the probability distributions be transformed into standard normalized

random variables (Equations A.19) and then the limit state function is mapped into a normalized coordinate system (Equation A.22).

$$g(X) = g(\{x_1, x_2, \dots x_n\}^T) = 0 \tag{A.18}$$

$$u_i = \frac{x_i - \mu_{x_i}}{\sigma_{x_i}} \tag{A.19}$$

$$U = \{u_1, u_2, \dots u_n\}^T \tag{A.20}$$

In this way, the CF becomes the solution to a constrained optimization in a standardized normal space. The solution of U represents the MPP.

$$\text{Minimize: } CF(U) = (U^T U)^{1/2} \tag{A.21}$$

$$\text{Subject to: } g(U) = 0 \tag{A.22}$$

Iterative solution methods are then employed using first or second order Taylor series expansions.

$$\tilde{g}(U) = g(U^*) + \nabla g(U^*)^T (U - U^*) + \frac{1}{2}(U - U^*)^T \nabla^2 g(U^*)(U - U^*) \tag{A.23}$$

This approach to QMU CF problem solving is different from methods in current practice. The method outlined requires an analytic expression for the limit state function and the variables therein represent distributions. Given the extra complications that arise from distributions that are not Gaussian it is difficult to imagine this method applied to the solution of limit state functions modeled with complex numerical simulations. The more important point is that with current QMU methods, the calculated CF is likely not the MPP solution.

Reliability and Confidence Levels

Predicted Reliability

The probability of failure can be related to the CF using Equation A.24. The reliability is R, and Φ is the cumulative density function.

$$R = \Phi(CF) \tag{A.24}$$

The two graphs in Figure A.1 show reliability versus CF for a Gaussian normal distribution. In the left graphic, a 50/50 reliability corresponds to a zero CF and zero M. As CF gets above 3, the reliability is expressed in terms of how many 9's follow the decimal place. The right figure plots with a logarithmic reliability axis showing the quantity of 9's. For example, at CF=6, the reliability is 0.9_9 or 0.999999999.

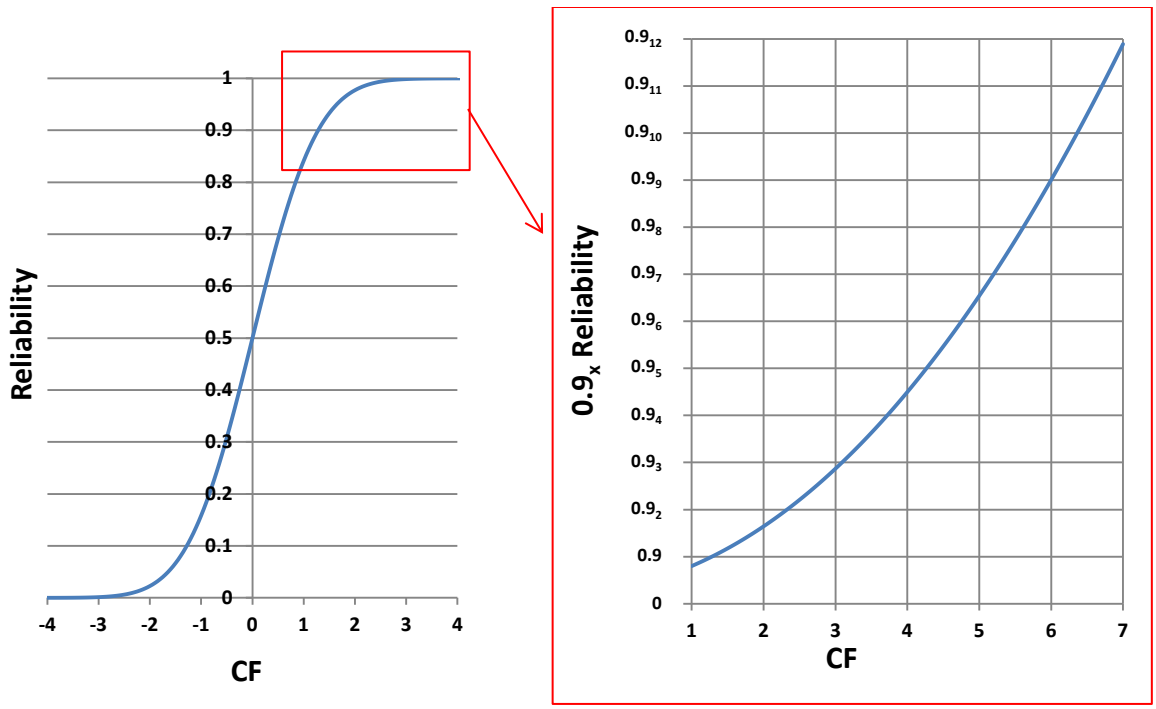

Figure A.1: Predicted reliability versus CF

The figure caption above labels the curve predicted reliability because for most QMU assessments, the M and U is estimated and is not based on actual statistical test data. When performing reliability testing, the sample number size is the key parameter to defining the confidence level of the statistical probability.

Mean Confidence Interval

When estimating the mean of a sample population with an assumed Gaussian normal distribution and unknown standard deviation, it is possible to define a bounding range for the mean using the t statistic. In the lookup of t values, the confidence level is $\left(\frac{\alpha}{2}, 1\text{-}\frac{\alpha}{2}\right)$, and ν is the degrees of freedom, $n\text{-}1$. The true mean is μ. $\bar{X}$ is the measured mean and S is the measured standard deviation (Equation 7). Equation A.25 defines the mean confidence bounding interval.

$$\bar{X} - t_{\left(\frac{\alpha}{2},\nu\right)}\left(\frac{S}{\sqrt{n}}\right) \leq \mu \leq \bar{X} + t_{\left(1-\frac{\alpha}{2},\nu\right)}\left(\frac{S}{\sqrt{n}}\right) \tag{A.25}$$

For the threshold distribution in Figure A.1, assuming a sample size of 20, the 95% (α = 0.05) mean confidence level interval is (2.53, 3.47). Table A.2 below is a small excerpt of t data at 95 and 99% confidence levels. It is worthwhile to note that past 30 sample points, the t data does not vary significantly, so that the bounds are narrowing mostly as the result of the $\frac{1}{\sqrt{n}}$ term. Figure A.2 shows the trending of the mean with sample size for two confidence levels.

$$\left(3 - 2.09 \cdot \left(\frac{1}{\sqrt{20}}\right), 3 + 2.09 \cdot \left(\frac{1}{\sqrt{20}}\right)\right) = (2.53, 3.47) \tag{A.26}$$

Table A.2: t Statistic Data

ν n-1 n_1+n_2-2	$\left(1-\frac{\alpha}{2}\right)=$ 0.995 $\frac{\alpha}{2}=0.005$	$\left(1-\frac{\alpha}{2}\right)=$ 0.975 $\frac{\alpha}{2}=0.025$
4	4.604	2.776
9	3.250	2.262
19	2.861	2.093
30	2.750	2.042
60	2.660	2.000
100	2.626	1.984

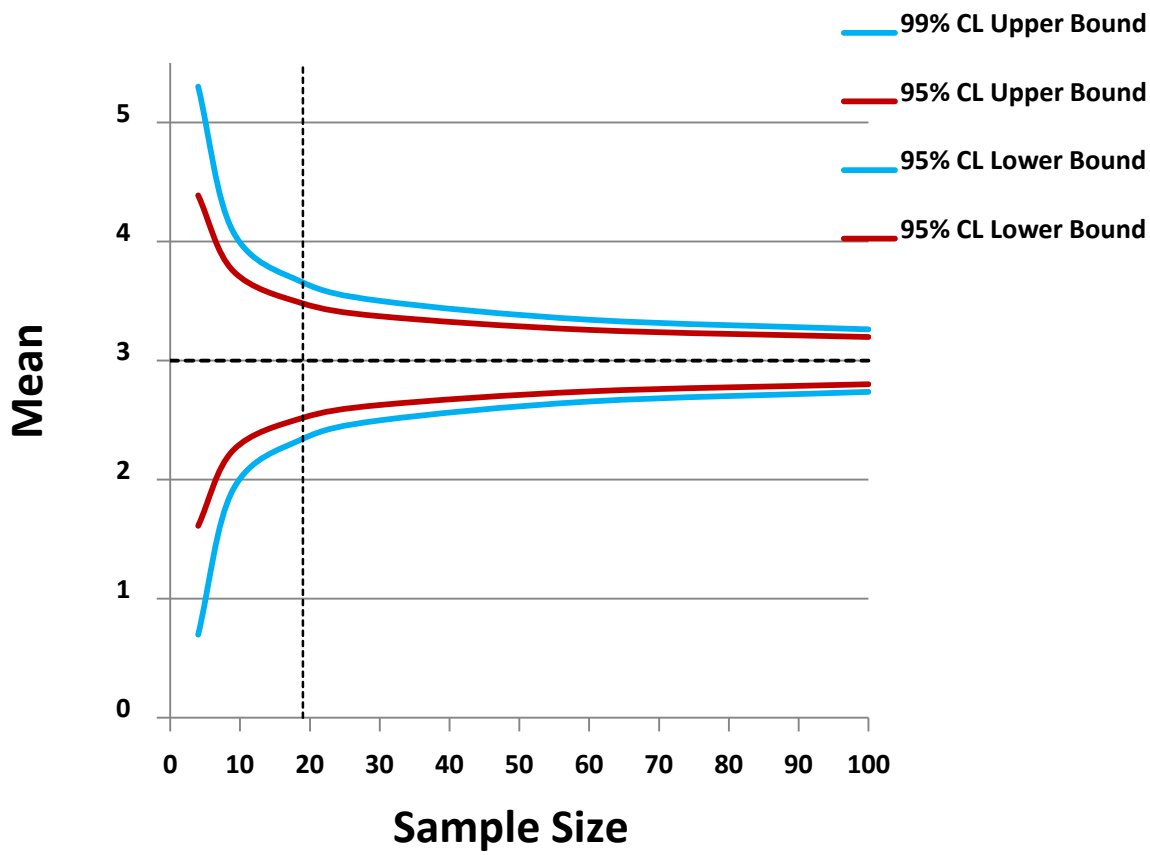

Figure A.2: Mean Interval curves versus Confidence Levels (CL) and sample size

Variance Confidence Interval

Bounding variances, based on a confidence level interval $\left(\frac{\alpha}{2}, 1\text{-}\frac{\alpha}{2}\right)$, for a sample size *n* are calculated using Equation A.26. $\mathcal{X}^2$ is a statistical function. When the standard deviation σ is estimated from a sample, it is denoted with an *s*. Table A.3 below is a small excerpt of $\mathcal{X}^2$ data for 95 and 99% confidence levels.

$$\frac{(n-1)s^2}{\mathcal{X}^2_{\frac{\alpha}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\mathcal{X}^2_{1-\frac{\alpha}{2}}(n-1)} \qquad \text{(A.26)}$$

The confidence interval for the standard deviation is the square root of the variance. For the nominal distribution shown in Figure 1 with a standard deviation of 1, the 95% confidence interval for the standard deviation with a sample size of 20 is between 0.76 and 1.46. Figure A.3 shows the variance trending of the example confidence interval with sample size for two levels of confidence α. Note the asymmetric differences in the upper and lower bounds and that sample size has a more pronounced effect on the upper confidence level bounds.

$$\left(\sqrt{\frac{19 \cdot 1}{32.85}}, \sqrt{\frac{19 \cdot 1}{8.91}}\right) = (0.76, 1.46) \tag{A.27}$$

Table A.3: Chi Squared statistical data

ν *n-1*	$\left(1-\frac{\alpha}{2}\right) = 0.995$	$\left(1-\frac{\alpha}{2}\right) = 0.975$	$\frac{\alpha}{2} = 0.025$	$\frac{\alpha}{2} = 0.005$
4	0.207	0.484	11.143	14.860
9	1.735	2.700	19.023	23.589
19	6.844	8.907	32.852	38.582
30	13.787	16.791	46.979	53.672
50	27.962	32.348	71.424	79.512
100	67.312	74.216	129.565	140.179

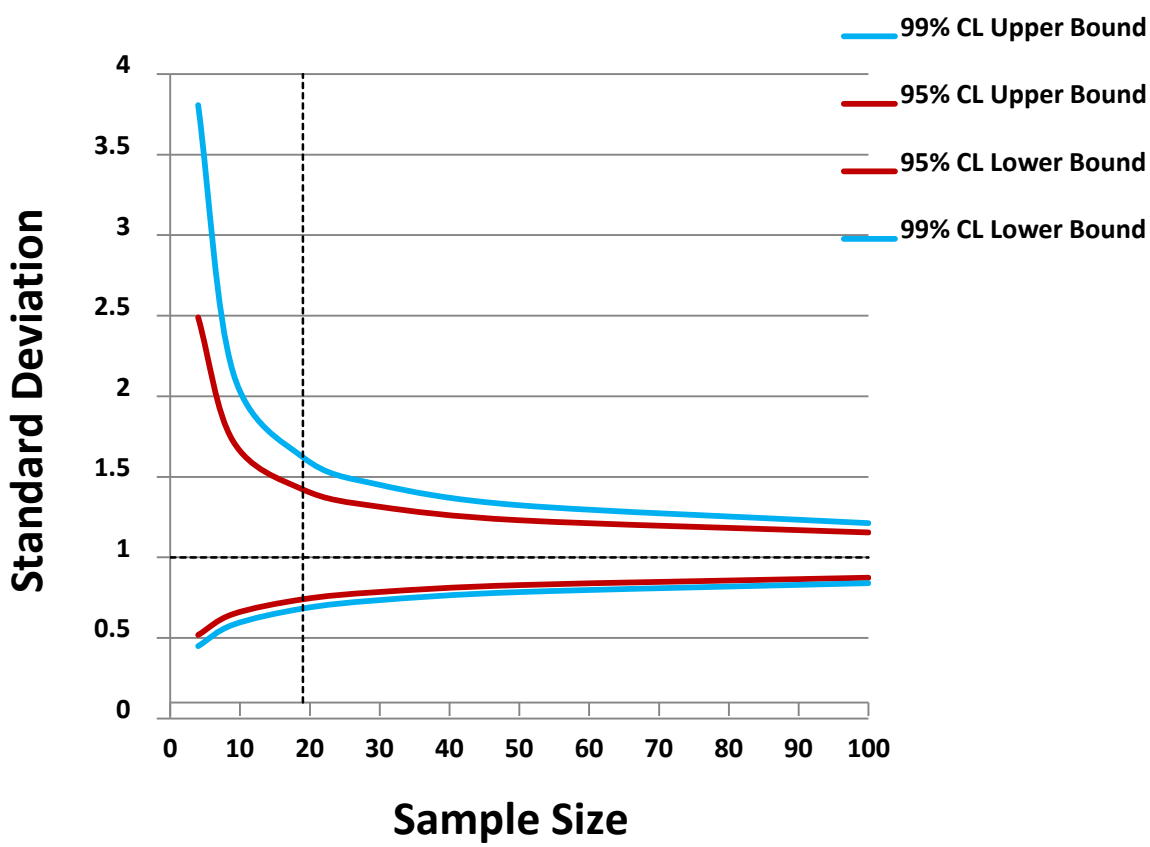

Figure A.3: Standard deviation interval curves versus Confidence Levels (CL) and sample size

Margin Confidence Interval

The statistical method used to test if two independent, normally distributed samples are the same is to compute a confidence interval on the difference between their means. This method is testing if two independent distributions have a statistical difference. For example, are the heights of men and women statistically different? Stated differently, is there a positive margin between the heights of men and women. If the lower and upper bounds straddle zero, then the hypothesis is false at a specified confidence level.

There are methods for large and small sample sizes and the solution shown in Equation A.28 is for small sample sizes of n_1 and n_2. The measured margin $\bar{M}$ equals the difference in measured means $\bar{X}_1$ and $\bar{X}_2$. A key assumption is that the measured standard deviations, s_1 and s_2, for the two distributions are similar in magnitude. A single variance is used to calculate the mean difference confidence interval so the variance is called a pooled variance s_p (Equations A.28 and A.29). Likewise, degrees of freedom (Equation A.30) used in the t statics data is also pooled.

$$\bar{M} - t_{\left(\frac{\alpha}{2},\nu\right)} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq (\mu_T - \mu_N) \leq \bar{M} + t_{\left(1-\frac{\alpha}{2},\nu\right)} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{A.28}$$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\nu}} \tag{A.29}$$

$$\nu = n_1 + n_2 - 2 \tag{A.30}$$

Using the nominal distributions shown in Figure 1 with margin M equal to 5, and standard deviations of 1, the 95% confidence interval for the margin with sample sizes of 20 and 12 is between 4.00 and 6.00. Like the t distributions for the mean confidence interval plot of Figure 7, similar t distribution bounding trends occur for the margin confidence interval.

$$\left(5 - 2.75 \cdot 1 \cdot \sqrt{\frac{1}{20} + \frac{1}{12}}, 5 + 2.75 \cdot 1 \cdot \sqrt{\frac{1}{20} + \frac{1}{12}}\right) = (4.00, 6.00) \tag{A.31}$$

Safety Factors and Reliability Design

Designing by Safety Factors (SF) is the most common mode by which engineers employ metrics to ensure that designs do not fail [35]. Designing for reliability requires extra steps that involve greater knowledge of the loading and failure distribution. This is especially true for small population designs where reliability cannot be tested or inspected [36].

Consider the case of a stressed part where the material failure strength exceeds the nominal stress by some ratio greater than one. Let the *SF* equal the ratio of the means of threshold strength to nominal stress

$$SF = \frac{\mu_T}{\mu_N}. \tag{A.32}$$

Next, the Confidence Factor equation (4) is rearranged in terms of the safety factor above and coefficients of variation, equation (9) [4].

$$CF = \frac{SF-1}{\sqrt{\eta_N^2+SF^2\eta_T^2}} \tag{A.33}$$

From inspection of Equation A.33, the sensitivity of the CF equally depends on the SF and the coefficients of variation η_N and η_T. In Figures A.4 and A.5 below, the CF is converted to design reliability (A.24), and η_T is assumed to be 0.05 and 0.15 so that two plots are created showing variations in reliability, *SF* and η_N. The previous notation of the quantity of nines is used to denote reliability. For instance, $R = 0.9_5$ is the same as 0.99999. As a means to explain and compare these plots, consider that when both η_N and η_T equal 0.05, there is 0.9_7 reliability for a SF of just 1.5 (Figure A.4). When η_N and η_T variations increase to 0.15 each, reliability drops to 0.9_5 and the SF must increase to 3.0 (Figure A.5). The singular increase of η_T from 0.05 to 0.15 greatly spreads the reliability curves apart.

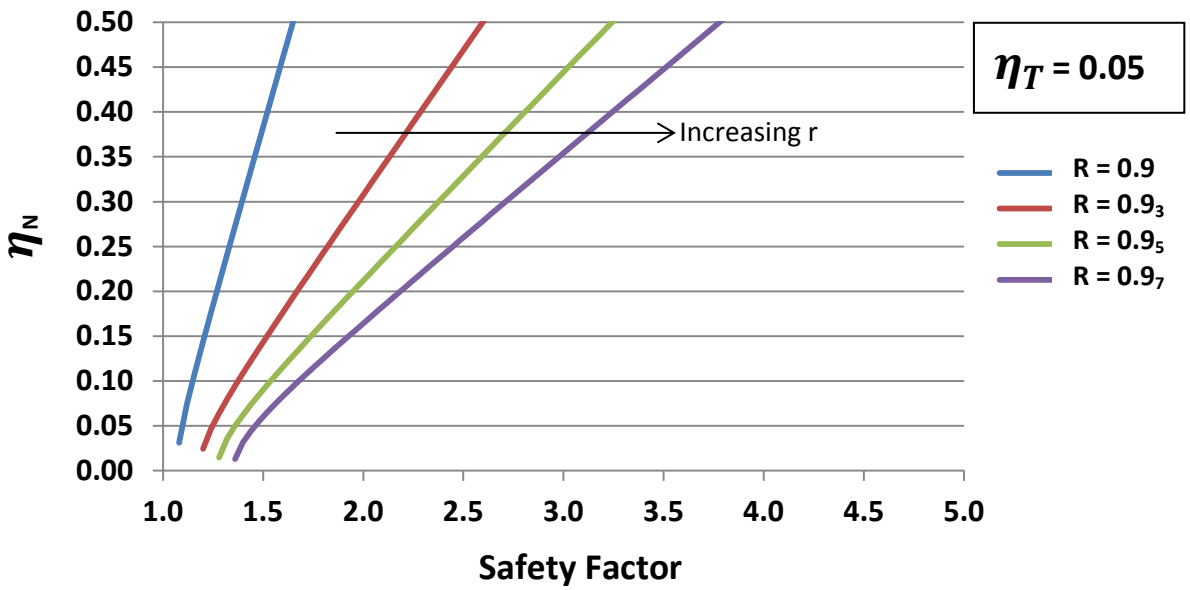

Figure A.4: Reliability, Safety Factors and the coefficient of Nominal variation for a 0.05 coefficient of Threshold variation

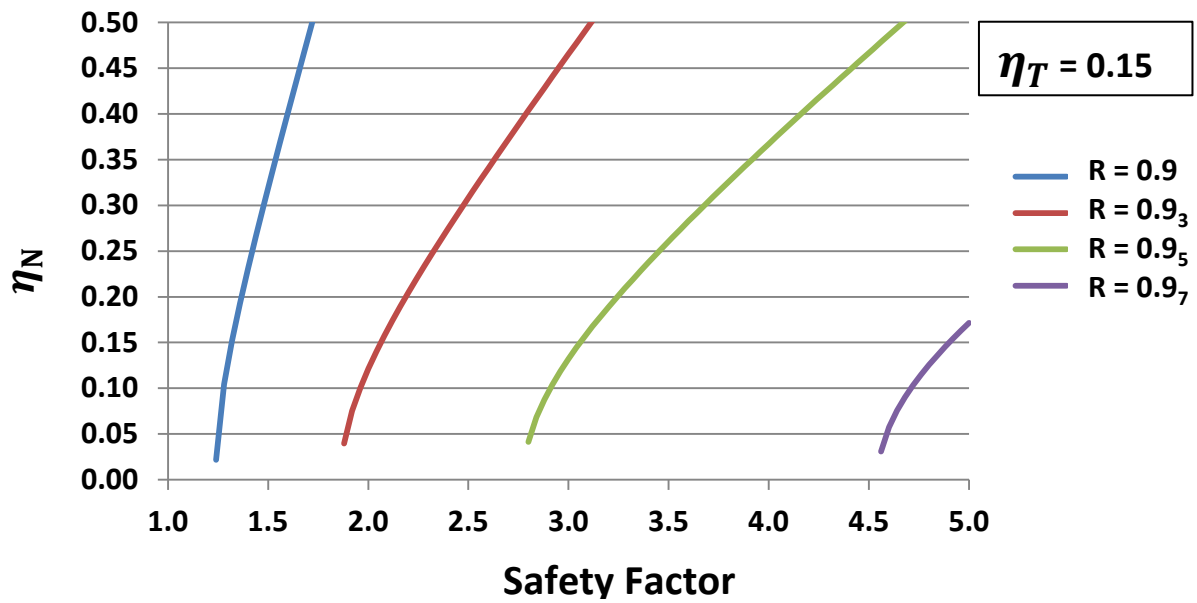

Figure A.5: Reliability, Safety Factors and the coefficient of Nominal variation for a 0.15 coefficient of Threshold variation

This examination of *SF* and CF shows what is intuitively obvious to scientists and engineers: variations in parameters that affect *SF* and CF are significant and they will have large impacts to design reliability. Designing for reliability requires consideration of coefficients of variability. When η_N and η_T are small (close to 0.05), then very high reliability is achieved and reliability is relatively insensitive to *SF*. It is when variations are larger that significantly large *SF*s are needed to achieve equivalent reliability.

Sandia K Factors

The development of a k-factor design approach at SNL traces to a 1963 monograph by D.B. Owen [37]. The monograph uses a simplified performance margin to define reliability. The margin is the distance between the mean of a nominal distribution to a constant failure threshold level *L* (Figure A.6). The figure shows a yellow region of failure to the left of the threshold value *L*. The probability is obtained from a one-sided t distribution Table A.1 [9]. From this direct approach to failure prediction, SNL defines a methodology whereby a relationship is established between k-factors (measure of reliability), testing sample sizes and confidence levels. The caution to note in this approach is the use of a singular threshold level *L*. This threshold value should be based on reliability and confidence levels that correspond to the required nominal performance reliability and confidence level.

The k-factor (Equation A.35) is the reliability measure that the nominal performance distribution exceeds a failure threshold. It is similar to the CF, with the major difference that a threshold constant replaces a threshold distribution. There is no threshold distribution tail to intersect the nominal distribution tail.

$$k = \frac{\mu_N - L}{\sigma_N} \tag{A.35}$$

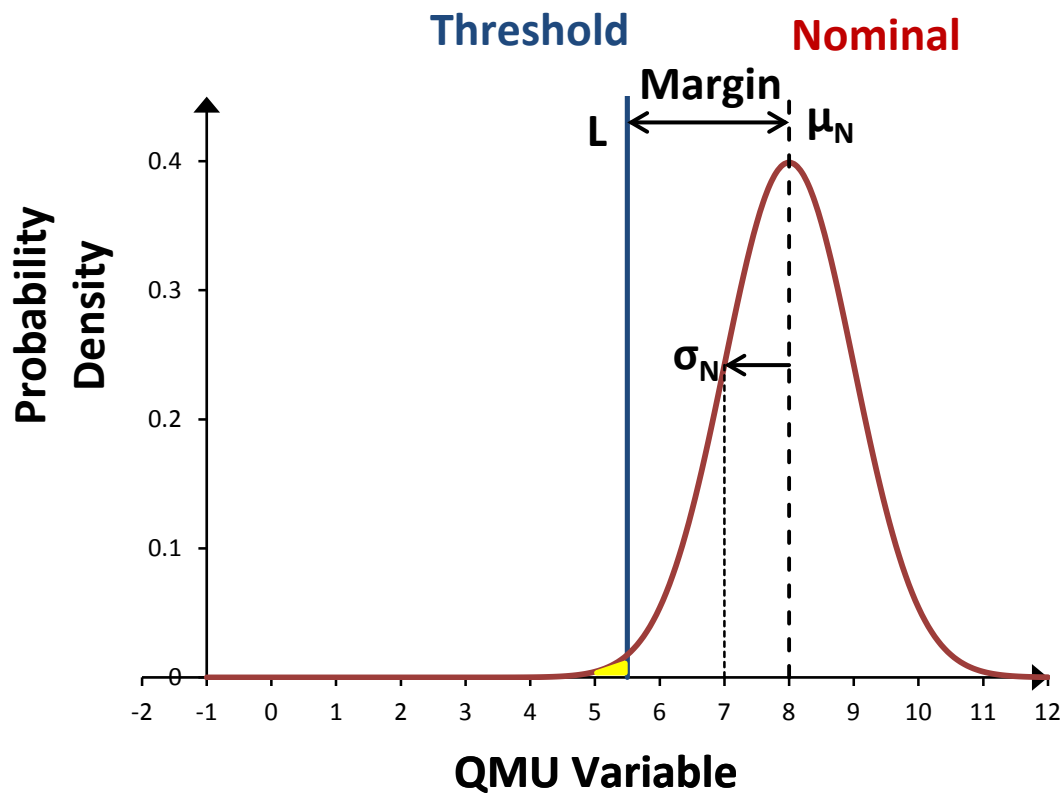

Figure A.6: Design Margin in a K-Factor analysis

From the previous confidence limit example plots, the interval limits are asymptotic at large *n*. Table A.4 shows the lower 90% confidence level bound k values versus sample size and reliability [37, 38]. For infinite sample sizes, the table is providing a minimum lower bound on the k value for an assumed confidence level (bottom row).

Table A.4: Lower Bound k factors[2] corresponding to 90% confidence level at given reliabilities

Sample Size n	Reliability 0.95	Reliability 0.99	Reliability 0.999
5	3.5987	4.9528	6.4951
10	2.6019	3.5818	4.6966
15	2.3443	3.2352	4.2470
20	2.2172	3.0661	4.0289
30	2.0850	2.8917	3.8051
200	1.7941	2.5152	3.3269
∞	1.6452	2.3268	3.0905

A look down the columns of Table A.4 shows a powerful relationship, useful for design purposes. The lower confidence bound for k is proportional to $1/\sqrt{n}$ (Figure A.7). To make use of this relationship, the

[2] The k factors have formula approximations [38, 39].

lower bound k factor for an infinite amount of test data, k_∞ is utilized to define a k Margin, Equation A.36. The observed k factor is calculated from test data. With sample size *n*, an observed nominal mean and standard deviation (Equations 7 and 8) are calculated and then used in Equation A.35 to calculate k_{Obs}. From a designers' perspective, the as-designed k factor, k_{Des} should be sufficiently large so that when tested with n samples, the observed k factor is larger than k_∞. Thus, k_∞ is selected based on design requirements. As an example using Table A.4, if there is a design requirement for 0.99 reliability at a 90% confidence level, then k_∞ is 2.3268. Theoretically, making a design with k equal to 2.3268 would be sufficient, but then it might be prohibitively expensive (infinite testing) to prove that the reliability requirement will be met. On the other hand, designing to k_{Des} equal to 3.6 establishes a k Margin of 1.2732 (3.6-2.3268). If the design is tested with 15 samples and the k_{obs} is greater than 3.2352, then the design meets the reliability and confidence requirements stated above.

Furthermore, three cells in Table A.4 are highlighted blue to show that with a design that has k_{Des} equal to 3.6, three different statistical statements can be made. With 5 tests and k_{Obs} > 3.5987, the design is 95.0% reliable at a 90% Confidence Level. With 15 tests and k_{Obs} > 3.2352 the design is 99.0% reliable at a 90% Confidence Level. Finally, with 200 tests and k_{Obs} > 3.3269 the design is 99.9% reliable at a 90% Confidence Level. There is an important risk versus cost tradeoff in this example. Testing quantities are usually the cost driver (building and expending hardware). The risk is making a successful design to a specification of k_{Des} equal to 3.6. Designing in a k margin mitigates design uncertainty so that fewer tests are necessary to demonstrate reliability and confidence level requirements.

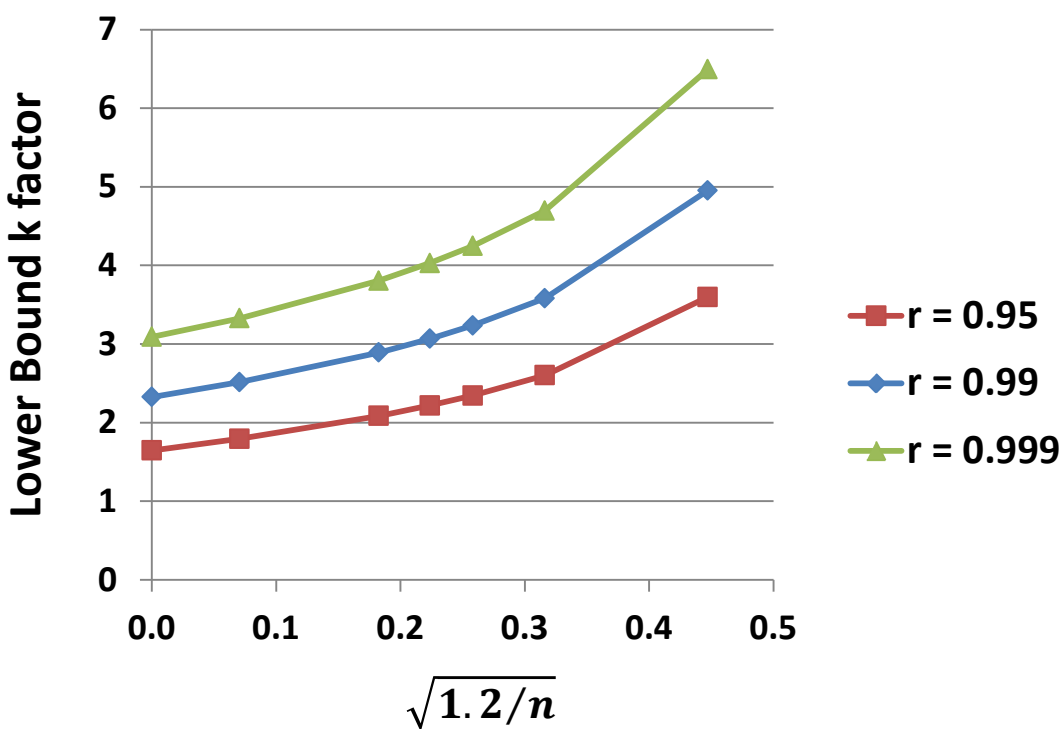

Figure A.7: Lower bound k factors for reliability at a 90% Confidence Level versus sample size n

$$k_{Des} - k_\infty \geq k_Margin = k_{Obs} - k_\infty \tag{A.36}$$

A bounding formula for the number of tests necessary to demonstrate reliability is derived from the the $1/\sqrt{n}$ trend to the statistical data of Table A.4 and the k_Margin (Equation A.37). With rearrangement, Equation A.38 defines the number of tests necessary show that a design meets reliability and confidence level requirements. Because this formula, like Table A.4 represents a lower bound k factor, the number of tests should always be rounded up to the next highest integer. The formula is also useful because it shows that for small k_Margin designs, sample sizes will have to be large. Figure A.8 plots the relationships of Equation A.38. The proportionality constant 1.2 fits the data and is applicable for the problem types (similar Confidence Levels and Reliabilities).

$$\sqrt{\frac{1.2}{n}} = \frac{k_{Obs} - k_{\infty}}{K_{Obs}} = \frac{k_Margin}{K_{Obs}} \tag{A.37}$$

$$n = 1.2\left(\frac{k_{Obs}}{k_Margin}\right)^2 \tag{A.38}$$

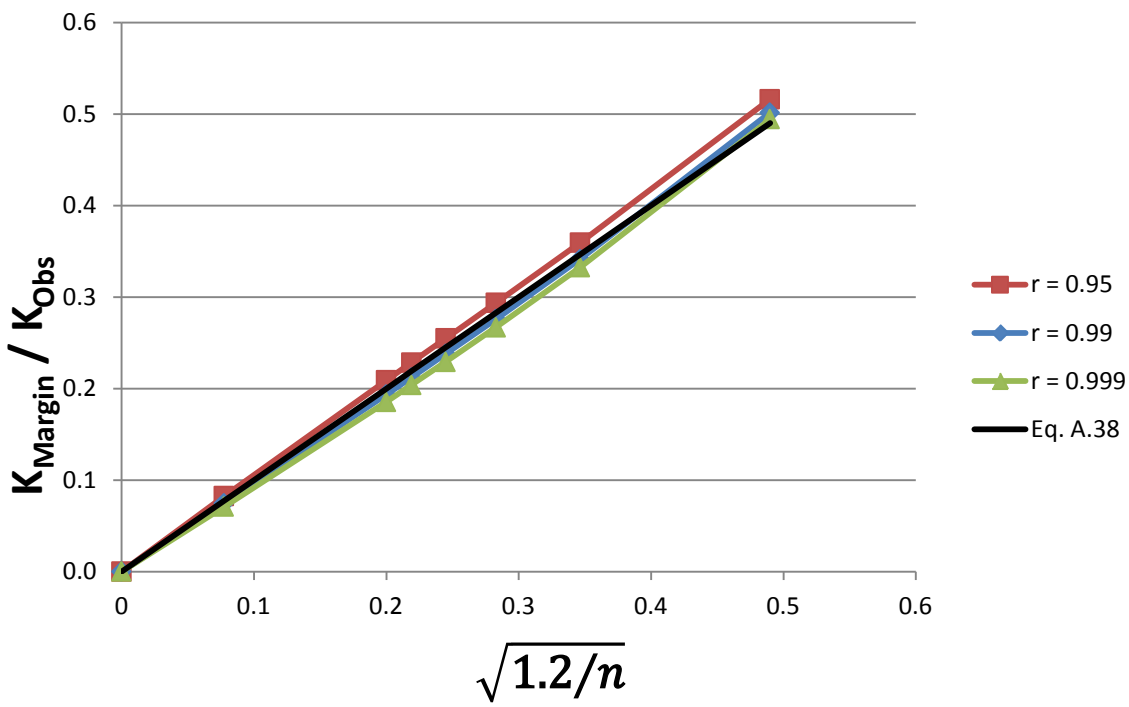

Figure A.8: k_margin versus sample size at the 90% Confidence Level

As a final k factors observation, the method is powerful because it provides a rational way to predict reliability and establish a confidence level plus significantly reduce test quantities. This is only possible because it is predicated on predictable designs with measurable k factors. In essence, designing a large performance margin ensures a large k factor and that results in high reliability. It is possible to reduce testing quantities with known high margin designs.

Measured Reliability

The previous section utilizes design based k factors and small sample size testing to predict and observe/measure reliability. In this section, a purely data driven approach is taken to measure and prove reliability. For component reliabilities where there is a presumption of low failure rate $\hat{P}$, with a discrete binomial distribution, and a large sample size n, a conservative Upper Confidence Limit (UCL) *C*, estimate can be made using the Maximum Likelihood Estimate (MLE) (Equation A.39) [40]. The reliability *R* equals *1-*$\hat{P}$, so the Reliability can be redefined and simplified in terms of confidence *C* (Equation A.39). A plot of failure rate probability (Figure A.9) clearly shows how measuring and proving high reliability and high confidence demands large scale testing and even larger (5-10X) total populations. The testing requires random draws from the population and if the tested population is large relative to the total population (10-20%) then a hypergeometric distribution assumption is applicable [41].

$$\hat{P} = 1 - (1 - C)^{1/n} \tag{A.39}$$

$$R^n = 1 - C \tag{A.40}$$

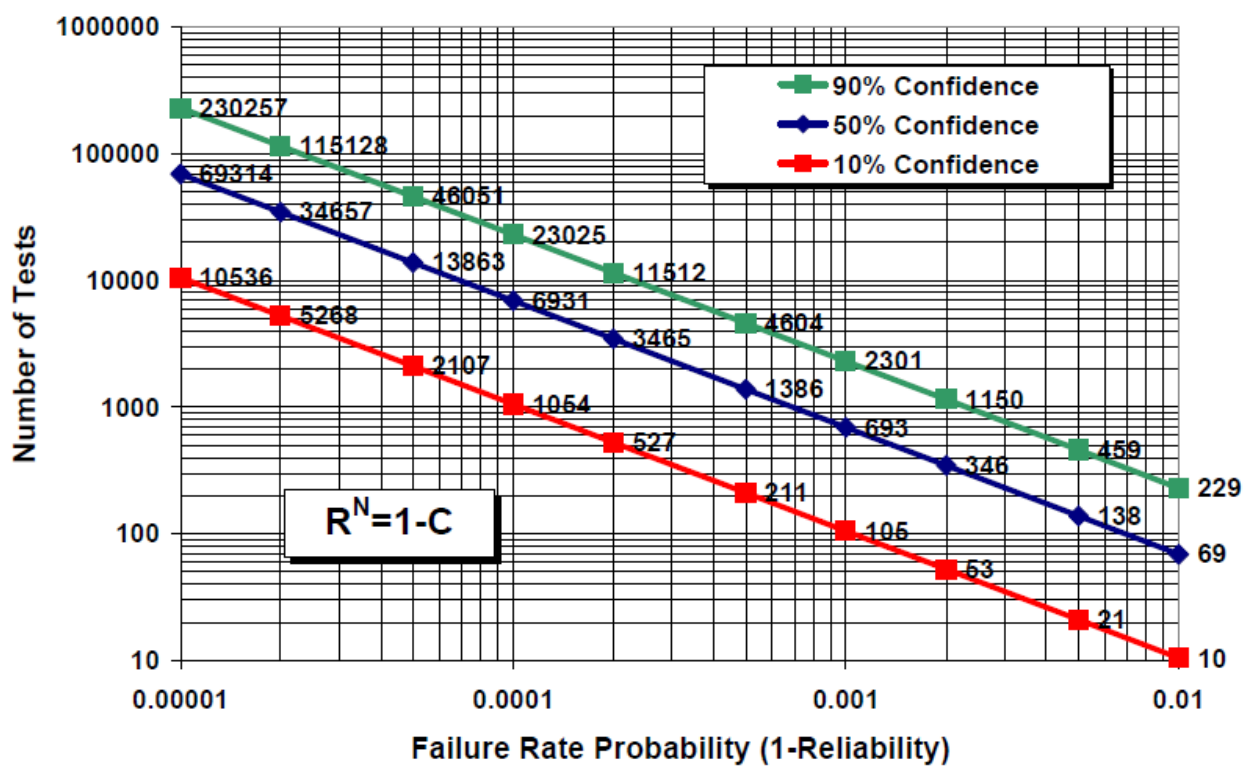

Figure A.9: Reliability testing requirements with no failures for various Confidence levels

Reliability is calculated at the 50% Confidence Level for nuclear weapons [20]. At a reliability level of *R* = 0.999, the required sample size is 693 tests without failure. If a failure occurs the reliability is calculated by dividing the number of failures by the number of tests. Note that calculating reliability at the 50% confidence level assumes ~0.7 failures have occurred. For example, if a failure occurred on the 693rd test for a component that required 0.999, the reliability would change from 0.999 to 0.9986, a decrease of

(only) 0.0004. However, in order to regain the required reliability, an additional 307 tests (1000 – 693) are necessary without failure.

Reliability was previously calculated at the 90% Confidence Level for nuclear weapons. For this same component, previously 2301 tests would have been required. Another view is that given 693 successful tests, the reliability is 0.999 at the 50% Confidence Level and 0.9967 at the 90% Confidence Level. The sample size decreases substantially as the Confidence level decreases. For example, given a 1000 sample size, the reliability is 99.99, 99.93 and 99.77% for Confidence levels of 10, 50 and 90% respectively.

LLNL[3] will sometimes utilize a combination of predicted and measured reliabilities to assess component reliability. It was previously mentioned that system reliability calculations use logic diagrams to construct formulas to calculate the overall system reliability. This same approach is taken at a component level using Equation A.41. The component reliability is the product of component QMU confidence factor reliability R_{CF} defined by Equation A.24 and the Dud failure rate defined in Equation A.39. The first term is predicted while the second term is measured. This is a conservative bounding approach that uses what we measure and know and what is grounded in our QMU prediction. Depending on the complexity of the component, more reliability terms can be added to represent the full component functionality. For this two-term reliability estimate, there are two important observations. The formula works best for high confidence designs (CF > 4) and it is conservative for designs with moderate to low confidence. For high CFs, the reliability R_{CF} has many 9s so the dud failure rate dominates. In this case, the failure rate depends on the quality of component fabrication. For the moderate to low CF cases, the dud failure rate may not be independent of failures due to design QMU. Hence, failures due to design might be double counted, resulting in a lower than actual reliability.

$$R = R_{CF}\left(1 - \hat{P}_{Dud}\right) \tag{A.41}$$

[3] Zoher Chiba at LLNL developed the QMU confidence factor with duds approach for reliability reporting

Appendix B

Surveillance Reliability and Confidence Sampling

The purpose of this Appendix is to provide a cursory explanation (not justification) for Nuclear Weapon Surveillance test quantities during the different weapon lifecycle phases. The NNSA surveillance requirements are defined in the Development and Production Manual [42]. Overall, surveillance is commonly referred to as the New Material Stockpile Evaluation Program (NMSEP), and it is divided into two phases, New Material and Stockpile Surveillance. New Material surveillance applies to new weapon systems. For stockpile systems that may undergo extensive retrofits or LEP re-acceptance and reuse, the program of surveillance for these units is referred to as Retrofit Evaluation System Test (REST).

During weapon production, the majority of the sampling activity consists of random Disassembly and Inspection (D&I) sampling followed by re-builds of the units. The sampling rate is defined by a 90/95 rule. The objective is to achieve a 90% probability *P* of sampling an effective unit rate *R* of 95% (5% defect rate), i.e., P/R. The second phase of NMSEP is the Stockpile Surveillance program and which begins 2 years after the start of production. In this phase, the sampling rate is defined by a 90/90/2 rule. Here again, the objective is to achieve a 90% probability of sampling an effective unit rate of 90% in a 2 year period.

Table B.1 [42] shows the sample sizes *n*. A cumulative sample quantity of $NT^{1/2}$ units is selected during *T* years of New Material production. *T* is in the units of years and N is the total production quantity plus the number of units rebuilt during the production period. What is remarkable about these sample sizes is that for the smaller populations, a large fraction of the population gets D&I surveillance. The first year of production surveillance does twice the surveillance amount as any subsequent year of new build surveillance. The major purpose of New Material surveillance is to detect and correct birth defects.

Table B.1: 90/95 Sample Sizes (90% chance that sample will allow detection of a 5% defect)

POPULATION	SAMPLE SIZE (n)	POPULATION	SAMPLE SIZE (n)	POPULATION	SAMPLE SIZE (n)
20-21	14	53-56	25	156-177	36
22-23	15	57-62	26	178-205	37
24-25	16	63-67	27	206-241	38
26-28	17	68-74	28	242-288	39
29-31	18	75-81	29	289-355	40
32-33	19	82-89	30	356-457	41
34-37	20	90-99	31	458-629	42
38-40	21	100-110	32	630-982	43
41-44	22	111-122	33	983-2130	44
45-47	23	123-137	34	2131-	45
48-52	24	138-155	35		

Table B.2 [42] shows the sample sizes *n* over a 2 year period based on a 90/90/2 rule. The surveillance activities include D&Is, destructive and nondestructive laboratory evaluations, and flight testing. What is again remarkable about Table B.2 is that for small populations, a large fraction of the population is tested. For the smallest population, one quarter of the stockpile would be disassembled and assembled each year? In practice, the warhead populations tend to the larger population sizes

Table B.2: 90/90/2 Sample Sizes (90% chance that sample will allow detection of a 10% defect in two years)

POPULATION	SAMPLE SIZE (n)	POPULATION	SAMPLE SIZE (n)	POPULATION	SAMPLE SIZE (n)
20-23	11	41-48	15	103-145	19
24-27	12	49-60	16	146-234	20
28-33	13	61-77	17	235-531	21
34-40	14	78-102	18	532-	22

Equation B.1 defines the P/R formula which is the basis for NNSA sample quantities. *N* is the total population size and n is the sample size. In the use of this formula, only the total population *N* is relatively fixed. The true weapon reliability *R* is an unknown defect fraction that changes over the weapon lifecycle. An objective in the surveillance sampling strategy is to detect aging related changes and to predict weapon reliability as it may change. The probability of detecting the defect fraction will in the end depend most on the sample size *n*.

$$1 - P = \prod_{i=1}^{n} \frac{(RN-i)}{(N-i+1)} \tag{B.1}$$

It is the authors opinion that the use of a sampling period "/2" is misleading because it is an accumulative probability of detecting defects. When viewed on solely on a yearly basis, the probability of detection drops significantly. It easy to show that if the sampling quantity is reduced, while overcompensating with an increased detection period that the detection probability increases. This counter-intuitive result is confusing and sends out the wrong signals to external customers.

The P/R formula has 4 independent inputs *(n, N, P, R)* with a single constraint of $n < N$. Figure B.1 plots probability versus sample size with a fixed population of 200. Figure B.2 is the same but with a population increased to 2000. For reliabilities near 90%, the two sets of results are quite similar. Only the plot of 99% reliability significantly changes. A "standard" weapon surveillance sample quantity is 22 units over a year period. From the figures, this corresponds to the 90/90/2 rule. At $n = 22$, that corresponds to 70% probability at $R = 0.95$, or 98% probability at $R = 0.85$. To achieve 95% probability at $R = 0.95$ requires a doubling of the sample rate. It is fairly concluded that 90/90/2 at $n = 22$ represents a balance in reliability and probability for a reasonable sample size.

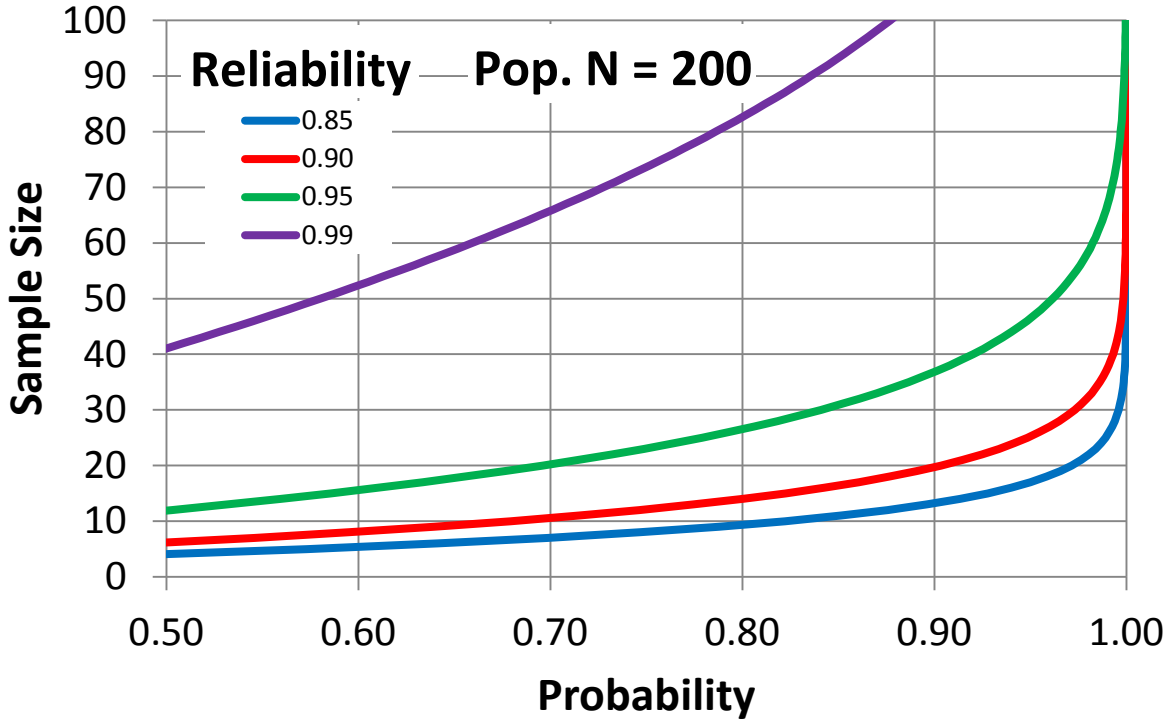

Figure B.1: Probability versus Sample Size for a small population

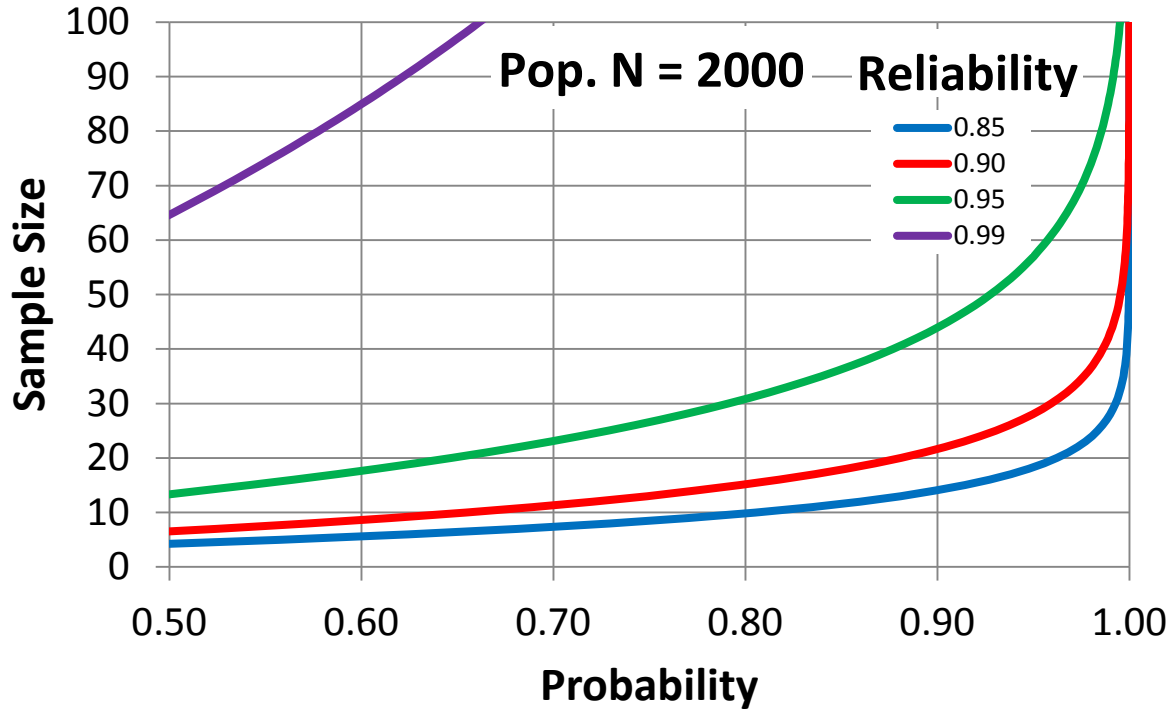

Figure B.2: Probability versus sample size for a large population

In the previous figures, the population size had a relatively small effect around 90/90. A different way to view the sample size is by the fraction of the total population sampled (Figure B.3). A key surveillance decision is what fraction of the total population should be tested over the total stockpile period? Per Figure B.3 and assuming 11 units per year lifetime, the maximum number of surveillance cycles is 18 and 45 for the blue and red curves respectively.

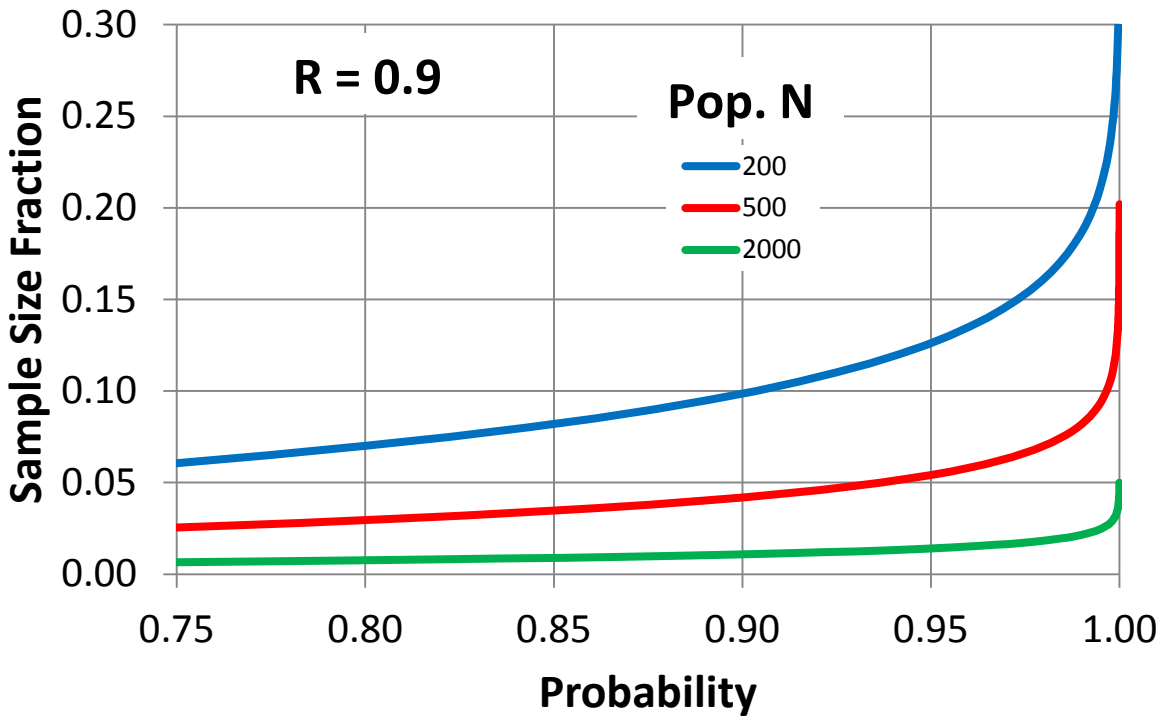

Figure B.3: Probability versus sample population fraction

Finally, it is helpful to understand the relationship between probability of detection and reliability inherent within the P/R formula. Figure B.4 shows that detection probabilities span from 0 to 100% over the short range of 80 to 100% reliabilities. The plots are entirely insensitive to population sizes between 200 and 2000.

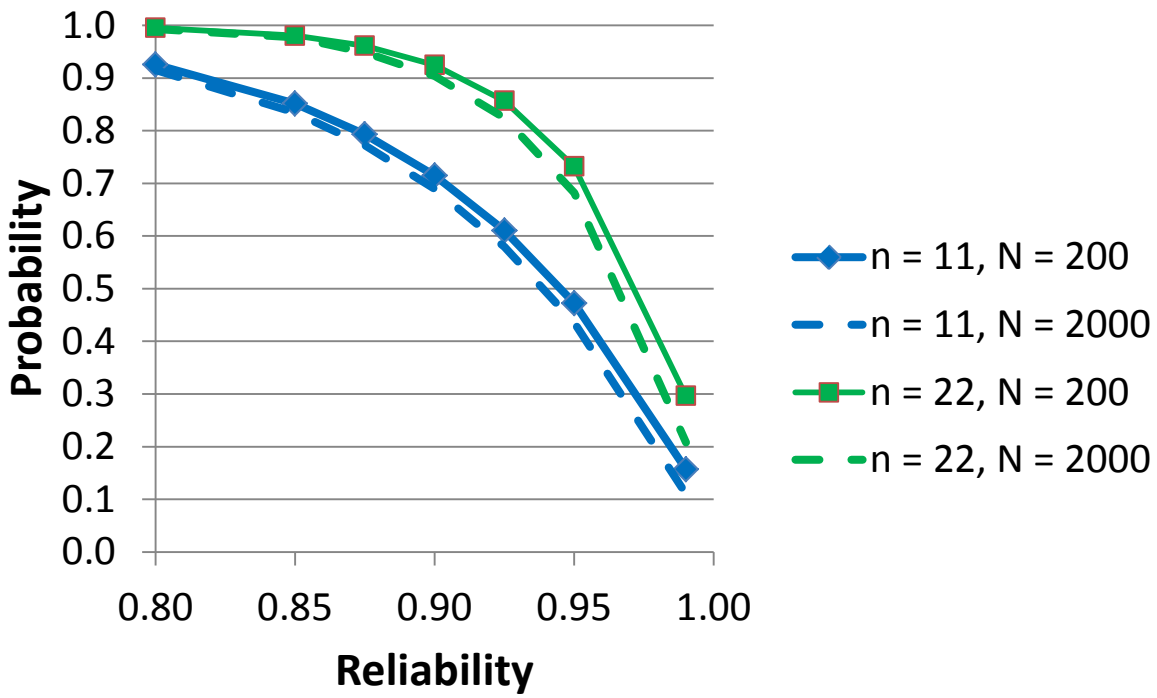

Figure B.4: Surveillance probability versus reliability

A common question asked in nuclear weapon surveillance is "What is the impact of doing less surveillance?" The response, especially if it is statistically based is usually less than satisfactory. Figure B.5 shows two prominent probability effects based on sample sizes and sampling periods. The solid green line represents "standard" surveillance: 22 units over a 2 year period yielding a 90% probability of detecting a 10% defect fraction. If the number of surveillance units decreases to 18 and 14 units, the corresponding probabilities (holding R fixed) decrease by 5% and 13% to 85% and 77% respectively. On the other hand, if the "standard" surveillance is restated as 11 units per year with a consistent 10% defect fraction, then the probability of detection in one year decreases to 69%. The dashed lines in the plot show the consequence of decreased detection periods. The purple curve shows the special case where the yearly surveillance rate is decreased to 9 units per year and the detection period is increased to 3 years. In this comparison to "standard" surveillance, the probability of defect detection increases from 90 to 94%, while doing less.

A remedy to the P/R inconsistency of surveillance periods is to always report the yearly detection rate and then to provide a cumulative probability of detection that spans the weapon lifetime.

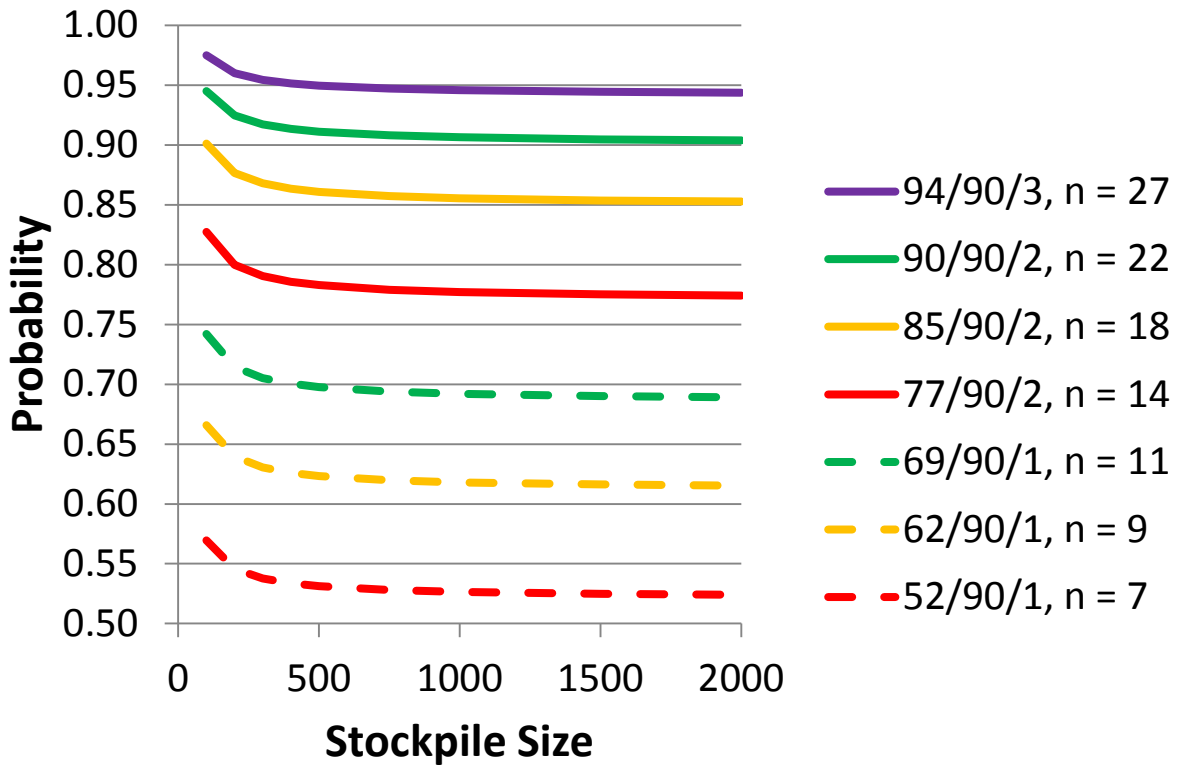

Figure B.5: Stockpile size versus probability

References

1. Goodwin, B.T., and Juzaitis, R.J., "National Certification Methodology for the Nuclear Weapons Stockpile", UCRL-TR-223486, Lawrence Livermore National Laboratory, Aug. 2006
2. Verdun, C., and Walter K., "A Better Method for Certifying the Nuclear Stockpile", Science and Technology Review, Lawrence Livermore National Laboratory, March 2004.
3. Ang, A.H. and Tang, W.H., Probability Concepts in Engineering Planning and Design, John Wiley & Sons Inc., 1975.
4. Kapur, K.C. and Lamberson, L.R., Reliability in Engineering Design, John Wiley & Sons, 1977.
5. Haugen, E.B., Probabilistic Mechanical Design, Wiley & Sons Inc., 1980
6. O'Conner, P. D. T., Practical Reliability Engineering, Second Edition, John Wiley & Sons, 1985.
7. Goldberg, B. and Verderaime, V., "Quasi-Static Probabilistic Structural Analyses Process and Criteria", NASA TP-1999-209038, Marshall Space Flight Center, Jan 1999.
8. Hayter, A.J., Probability and Statistics for Engineers and Scientists, Duxbury, Wadsworth Group, 2002
9. Ayyub, B.M., and MCuen, R.H., Probability, Statistics, and Reliability for Engineers and Scientists, Chapman & Hill CRC Press LLC, 2003
10. Choi, S., Grandhi, R.V., and Canfield, R.A., Reliability-Based Structural Design, Springer-Verlag, 2007
11. Nowak, A.S., "Survey of Textbooks on Reliability and Structural Safety", The Structural Engineering Institute of the ASCE, Univ. of Mich., Aug 2004
12. Tait, N.R.S., "The use of probability in engineering design – an historical survey", Reliability Engineering and Systems Safety, 40 (1993) 119-132.
13. Pugsley, A. G. & Fairhorne, R. A., "Structural research in aeronautics", Aircraft Engineering, June (1939) 225-7.
14. Acar, E. and Haftka, R.T., "Reliability Based Aircraft Structural Design Pays Even with Limited Statistical Data", 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference May 006
15. Acar, E., Aircraft Structural Safety: Effects of Explicit and Implicit Safety Measures and Uncertainty Reduction Mechanisms, PhD Dissertation, Univ. of Florida, 2006.
16. "Structural Design and Test Factors for Spaceflight Hardware", NASA-STD-5001:, section 3. NASA, 2008.
17. "Strength and Life Assessment Requirements for Liquid Fueled Space Propulsion System Engines", NASA-STD-5012, NASA, 2006
18. Bierbaum, R.L., Cashen, J.J., Kerschen, T.J., Sjulin, J.M., and Wright, D.L., "DOE Nuclear Weapon Reliability Definition: History, Description, and Implementation", SAND99-8240, Sandia National Laboratory, Apr 1999.
19. Love, S.L., "A History of Stockpile Quality Assurance at Sandia Laboratories", SAND79-0696, Sandia National Laboratory, Feb 1980.
20. Wright, D.L., and Bierbaum, R.L., "Nuclear Weapon Reliability Guide Evaluation Methodology", SAND2002-8133, Sandia National Laboratory, 2002
21. Kane, R.J., "Component Ratings and Reliability Design", LLNL-TM-628223, Design Safety Standards Section 11.7 Lawrence Livermore National Laboratory, Mar 2013.
22. Montgomery, D. C., 2000, Design and Analysis of Experiments, 5th Ed., John Wiley, Hoboken, NJ.
23. "General Requirements 9900000", Issue AP, National Nuclear Security Agency, 2005
24. Oberkampf, W.L., DeLand, S.M., Rutherford, B.M., Diegert, K.V., and Alvin, K.F., "Error and Uncertainty in Modeling and Simulation", Reliability Engineering and System Safety, 75, Elsevier, 2002, pp 333-357

25. Oberkampf, W.L., DeLand, S.M., Rutherford, B.M., Diegert, K.V., and Alvin, K.F., “Estimation of Total Uncertainty in Modeling and Simulation”, SAND2000-0824, Sandia National Laboratory, Apr 2000
26. “Systems Engineering Fundamentals”, Systems Management College, Defense Acquisition University Press, January 2001
27. “Systems Engineering Handbook”, SP-2007-6105 Rev 1, NASA, Dec 2007
28. “Systems Engineering – Application and management of the systems engineering process”, IEEE Standard 1220-2005, 2007
29. RMI, Writing Effective Product Requirements, T060-A, National Nuclear Security Agency, 2010
30. Malakoff D. “Italian Quake Verdicts Rattle Researchers”, Science, Dec 21, 2012, pg. 1526.
31. Weiss, C., “Scientific Uncertainty in Advising and Advocacy”, Technology in Society 24 (2002) 375–386, 2002
32. “Hazard Analysis Reports for Nuclear Explosive Operations”, US Department of Energy, Washington D.C., DOE-NA-STD-3016-2006
33. Hasofer, A.M. and Lind, N.C., ”Exact and Invariant Second-Moment Code Format”, ASCE J. of the Engineering Mechanics Division, Feb. 1974
34. Parkinson, D.B., “Solution for Second M*omen*t Reliability Index”, ASCE J. of the Engineering Mechanics Division, Oct. 1978
35. Murray, R.C., “Safety Factors”, Lawrence Livermore National Laboratory Engineering Design Safety Standards Section 1.1 Rev. 3, June 2012
36. Aggarwal, P., “Structural Requirements for the Space Propulsion Engine Systems”, N20060007807, US Department of Commerce, National Technical Information Service, 2006
37. Owen, D.B., “Factors for One-Sided Tolerance Limits and for Variables Sampling Plans”, SCR-607, Sandia Corporation, Mar 1963.
38. Natrella, M.G., “Experimental Statistics”, National Bureau of Standards Handbook 91, Ch. 2-1, August 1, 1963, Reprinted October 1966.
39. Link, C.L., “An Equation for One-Sided Tolerance Limits for Normal Distributions”, US Dept. of Agriculture, Forest Service, Research Paper FPL 458, 1985.
40. Diegert, K.V. and Zurn, R.M., “Nuclear Weapon Reliability Guide 2nd Edition”, SAND93-704, Sandia National Laboratory, 2002
41. Norman L. J., Adrienne W.K. and Kotz S., Univariate Discrete Distributions, John Wiley & Sons., Third Edition, 2005,
42. “New Material and Stockpile Evaluation Test Program”, Development and Production Manual Chapter 8.1 Issue B, National Nuclear Security Agency, Dec. 2005