

From Model Validation to Reliability Assessment: Comments on Non-Deterministic Approaches (NDA)

Roger W. Logan

Copyright © 2009 SAE International

ABSTRACT

This paper is a discussion of topics presented by the author at the Panel Session “Evaluation of Studies on Non-Deterministic Approaches (NDA) for Complex Systems” at the 2007 SAE World Congress. The emphasis herein is on issues in conducting model based reliability assessments, by combining verification and validation (V&V) of the model with a process compatible with traditional reliability assessments conducted as part of Reliability Based Design Optimization (RBDO). Formulations are presented and simplified to isolate each of the terms in V&V, and make each term compatible with subsequent RBDO while still accommodating non-deterministic design situations. The paper concludes with overall issues regarding V&V and Reliability, and compares this combined method with some other methods in use in the community and discussed at a follow-on SAE panel in 2008.

1. INTRODUCTION

This narrative will focus on the linkage between computer models, test data (which is typically all too scarce), and reliability based design. The discussion is intended to help the community establish a link between Model V&V (including emerging national standards) and design/product Reliability assessments. It is typically the case that the very reason that we are using a model, instead of just test data, is because there is not enough experimental data, and/or that data is not at the same conditions desired for real world use. This indicates that the linkage of the model to an assessment of the reliability of the real world product will involve elements of NDA. Usually, what we see falls into one of two extremes:

1. The first extreme: Reliability is based on a frequentist assessment: Test data is plentiful, enabling a binomial or “x/N” style reliability assessment. In this situation, there are established methods for quoting a value for the confidence in this assessed reliability as well. The frequentist assessment may remain the most foolproof method for assessing reliability, but we can rarely afford the thousands of tests required to use this method to attain high reliability and confidence. Eventually, during product use, we may obtain this invaluable data, but that is too late for the design process.
2. The second extreme: Reliability is computed from results of a model, and this *unvalidated* model is used to compute all the terms in the Reliability Index (•). Even if there is some ad hoc assemblage of test data to support the intuition used in selecting the model, this test data is not quantitatively integrated and linked from the Model through V&V through Reliability. One of the ways we can usually flag such a model is to ask, “what is the assessed Confidence C, where $0\% \leq C \leq 100\%$, in the model?” If there is no answer, or an answer like “high confidence” or “we didn’t assess that”, it is usually because the model has not undergone V&V. The Reliability assessment may still be first-rate, and the Model may be excellent, but this is based on trust and experience, not on the use of a Validated Model. In some situations, this may be good enough, but we must go further when it is cost-effective to do so.

The goal of this discussion is to add a 3rd and preferred scenario that strikes the best balance between [1] and [2], as follows: Reliability is computed from the results of a model. That model has been validated over some domain of test data, so that a numerical Confidence $0\% \leq C \leq 100\%$ can be quoted to accompany any assessment that uses that model. If this cannot be done, we have failed to credibly connect the Model and the assessed product reliability. If it can be done, then we have successfully connected the Model, the expense of computing, and the expense of testing, to the reliability and confidence in our design and product. As by-product, we also have a quantitative means to defend an extra test or extra series of computer runs we might want to perform.

The benefits and dangers of using numerical models to assess product reliability can be summarized as follows:

*“Design provides Reliability.
Testing provides Confidence.
Our Model quantifies both.
That’s the scary part”*

The statement, an excerpt from the 2007 Panel Discussion, is a slight oversimplification. However, it is important to stress the distinction between reliability and confidence, and that the two are equally important and equally essential. One senior official responsible for assessments of our strategic defense forces recently reflected on the topic with a similar one-liner: “Reliability is useless without an associated Confidence”.

To state what is perhaps obvious, we can have a model of our design, and we can use numbers from the model to assess the Reliability at $R=98\%$. That means that our assessment is that if we put all of the hundreds or thousands of our product items in use, 98% of them would pass and 2% of them would fail, and our assessment is that we expect these 2% to fail.

We could also make such an assessment based on test data. Let’s say we test 6 items randomly sampled from our entire production run. All 6 “pass” the test. A simple binomial assessment says that we can claim 90% Reliability, but only at 50% Confidence (actually that is 50% 2-tailed Confidence). That means that we expect 90% of the production run to pass, and we expect 10% of the production run to fail, when we test them all, so $R=90\%$ is the assessed Reliability. We could have assessed over 99% Reliability and been equally credible; but we would have to state a (2-tailed) confidence of near $C_2 \sim 0\%$, or, assuming symmetry, a 1-tailed confidence $C_1 \sim 50\%$. The Confidence %C means, as far as the author’s intuition is able to express, the fraction of possible “universes” within which our assessed reliability holds. Since we don’t know which universe[s] we might be living in when the real product is asked to do its job, our Reliability assessment holds only

within the %C of the total possible universes. So if we assess $C_1 \sim 50\%$ Confidence, that means that in 50% of the possible universes (we might be in one of them), then for sure, at least 98% of our product will succeed and at most 2% will fail.

However, in the other 50% of the universes, (we might be in one of these, too), we have no idea what the reliability is. It might be $R=100\%$, or it might be $R=0\%$, or any number in between. That is why low assessed confidence is perhaps just as dangerous as low assessed reliability. To be fair, it is likely that the reason we do so well in the design community is that there is very good subjective judgment that the universe[s] where we have assessed the product reliability are in fact the right choice[s], so we are unlikely to encounter the other universe[s] where our assessment of reliability is completely wrong compared to reality. It is not new to associate reliability with a confidence level; this is typically done in consideration of a finite number of experimental “go or no-go” tests, often with an associated statement of “lower and upper [confidence] limits” on reliability.

Our goal in connecting a *Model*, and its value, to a reliability assessment, has to be to quantify the relation between confidence and reliability *bearing in mind our use of the model and its validation pedigree*, or else we will be unsuccessful in justifying the tests and computing effort we need to credibly assess the design. Or perhaps worse, we will be funded to build an elaborate millions-of-elements finite element model, whose sheer complexity compels a feeling of credibility, and we will then bet our company, or the lives of others, on this model. Since it is easier than ever to lose our design intuition when overwhelmed by millions of elements and colors, this situation can be more dangerous and misleading than if we had no model at all. At least then, we would realize that we are just betting on whatever meager data we have, along with the subjective, albeit often excellent, design expertise and choices of simple approximations made in the design assessment.

To encourage quantitative linkages between computer models and product design, we need a simple way to make a first-cut link between Models, V&V, and assessed confidence and reliability. It is hoped that this simple linkage can be done quickly (perhaps even in real time at the table of a design review), but that the linkage is fully compatible with far more complex methods for V&V and Reliability analysis. In an effort to offer that compatibility, the references include a cross-section of the literature along the way.

We can summarize the theme of this work by introducing the major points:

- Credible integration of models and design requires quantitative Model V&V.

- Quantitative Model Verification & Validation must quantify terms of four types as discussed below.
- Quantitative model V&V couples to the design to give bounds on reliability, i.e. the model assessed reliability as a function of confidence.
- The process described next is compatible with the newest V&V Standards, and couples readily into reliability based design.

2. SIMPLE BUT ILLUSTRATIVE NDA PROCEDURE TO LINK MODEL V&V TO PRODUCT RELIABILITY ASSESSMENT

Following is a simplified step by step methodology for a quantified linkage between Models, V&V, and Reliability in design. Detail, depth, and alternatives are provided by reference to existing standards, guides, and reviewed, published literature in this area. The example is relatively easy to follow, and is compatible with SAE^{45,48}, AIAA¹, and ASME^{5,6,7} guides and standards in this area. However, the quantitative path it illustrates from Model V&V to Reliability assessment has been greatly simplified for illustrative purposes.

Our goal will be to assess Reliability, using a combination of models (sometimes but not always finite element models) and data, with the data typically sparse and sometimes not of the pedigree we might desire. The challenge of cost is always present, so that while higher fidelity simulations and higher quality, more abundant validation-specific test data are always desirable, we must have a methodology that lets us say what we can say with limitations on our computational and test resources, and then make that methodology compatible with the case where we have all the computational and test resources we want.

The discussion that follows is simplified, but can be linked to more elaborate treatments of each topic already in existence. In other words, it is a procedure to go from a “potpourri of models and test data”, to

quantitative V&V, then on to a quantitative assessment of reliability at a specific confidence level.

2.1 SIMPLE FORMULA FOR RELIABILITY ASSUMING A NORMAL PDF

Consider (Fig. 1) a bolt with fixed load L and random variable Strength S , with mean μ_s .

Then the *Central Margin*

$$M = \mu_s - L \quad [1]$$

And, the *Central Factor of Safety*

$$FOS = \mu_s / L \quad [2]$$

Assume that L is known and fixed, and that the *mean* μ_s of Strength S is also known with 100% Confidence, and fixed, and that the random variable S has a known standard deviation σ_s .

If the distribution of S is normal, we assess the Reliability Index RI , or β , as:

$$\beta = (M/\sigma_s) \quad [3]$$

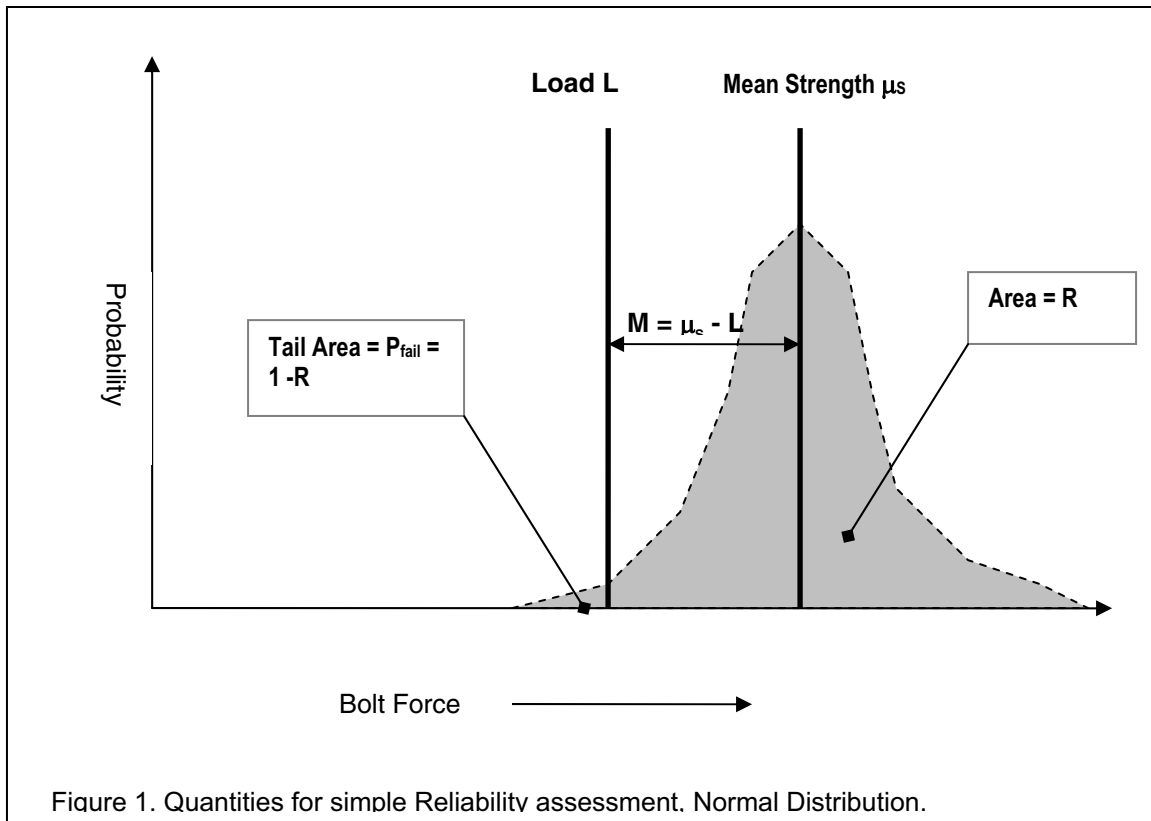
And we have Reliability R such that:

$$R = \Phi(\beta) \quad [4]$$

Table 1 shows the resulting 1-tail Reliability R vs. Reliability Index β :

Table 1. Reliability Index β and assessed Reliability R , Normal Distribution

β	R
0.0	.5000
1.0	.8413
2.0	.9772
3.0	.9986



2.2 SIMPLE FORMULA FOR RELIABILITY ASSUMING A UNIFORM PDF

Assume again that μ_s , the mean of S , is known at 100% Confidence and fixed, and that S has a standard deviation σ_s but with a *uniform* PDF (Fig. 2). Let the uniform PDF vary from a to b , where a is the lower limit and b is the upper limit.

Given this, the mean μ_s is equal to $(a+b)/2$ and the standard deviation σ_s is equal to $(b-a)/(2\sqrt{3})$. Now, the distribution of strength S varies uniformly from $\mu_s - \sigma_s\sqrt{3}$ to $\mu_s + \sigma_s\sqrt{3}$.

$$\text{Let } U = \sigma_s\sqrt{3} = (b-a)/2 \quad [5]$$

Then if we construct the Uniform Reliability Index, B_u , as the ratio of margin M (mean strength – mean load) normalized by half the range of the strength:

$$B_u = (M/U) \quad [6]$$

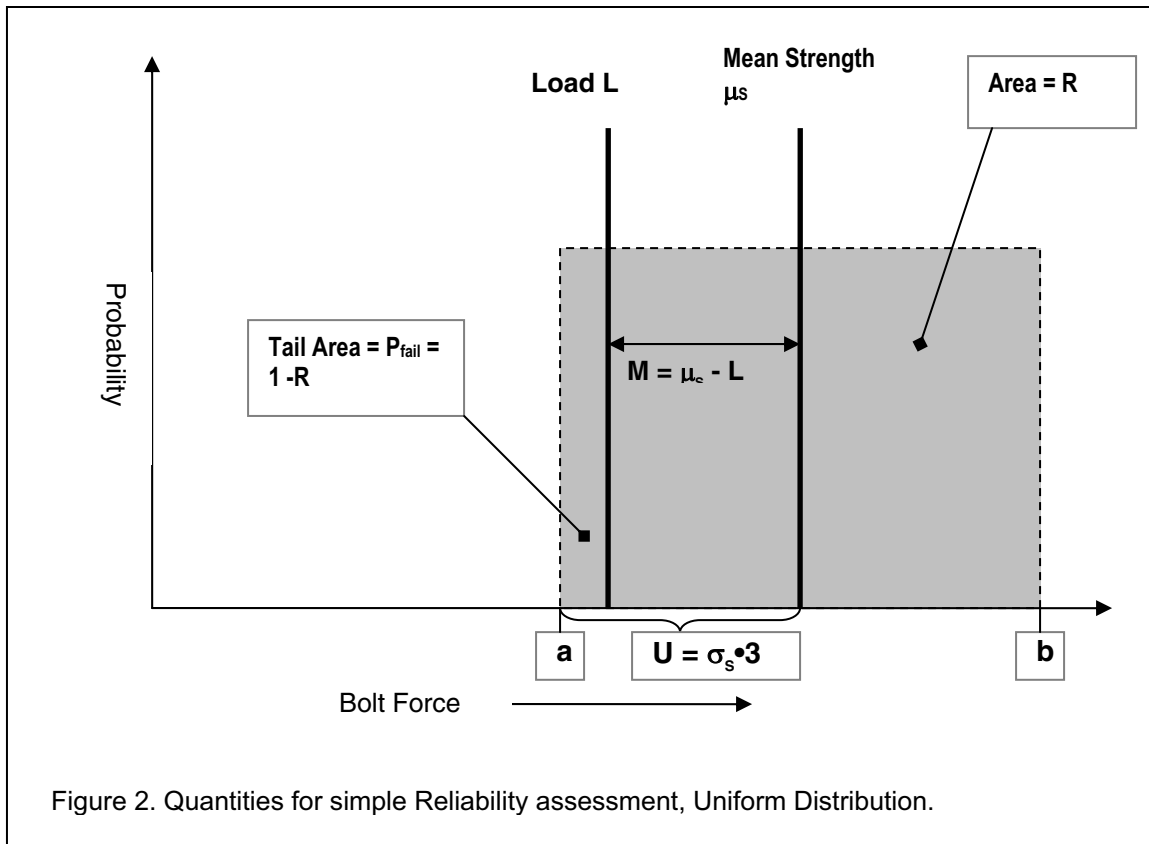
Note carefully that $B_u \cdot \beta$. We now have 1-tailed Reliability R_u such that:

$$R = R_u = 0.5 + 0.5 \cdot B_u \quad [7] \\ (0 \cdot R \cdot 1)$$

Table 2 shows the resulting 1-tail Reliability R_u vs. (Uniform) Reliability Index B_u :

Table 2. Reliability Index B_u and assessed Reliability R , Normal Distribution	
B_u	R_u
0.0	.5000
0.25	.6250
0.50	.7500
1.00	1.000

The above is an oversimplified depiction of reliability assessment; however it will suffice in order to make our points in linking V&V with Reliability.



2.3 DETAILS OF THE MODEL V&V TO RELIABILITY AT CONFIDENCE PROCESS

Consider the case above for the Normal Distribution:

$$\beta = (M/\sigma_s) \quad [3]$$

$$R = \Phi(\beta) \quad [4]$$

This looks simple enough, but we still must describe a procedure to obtain “M” and “ σ_s ”.

First, we will change notation from σ_s to σ_T , and denote this as the Standard (T=Total, from all contributors) Uncertainty of the Quantity Of Interest (QOI). The QOI is the focus of the Model V&V, and also the focus of our Reliability assessment equations above. We must now link our computer model results to these terms in Reliability, and we will do so using the products of Model Verification and Validation (V&V). This discussion borrows heavily from the language in the AIAA Guide for V&V¹, the new (Dec 2006) ASME V&V-10 Guide for V&V in Computational Structural Mechanics⁶, and the draft ASME V&V-20 Standard for V&V in Computational Fluid Dynamics and Computational Heat Transfer⁷.

Where possible, we follow the draft ASME V&V-20 nomenclature for Standard Uncertainty (that is, all types of relevant uncertainty terms “i” expressed as an estimate of the standard uncertainty, u_i (not corrected for small sample size). Each of these u_i terms is comprised of a bias term b_i and a random uncertainty term s_i (the standard deviation). To obtain the standard uncertainty

u_i , the terms b_i and s_i are added in quadrature. In practice, for design, we will later want to do two things:

1. “Dial out” the bias term b_i to calibrate our model to test data. Great caution is required here.
2. Assess our design at a numerical value of confidence, which may be different than the numerical value of confidence inherent in the standard uncertainty u_i (this is a function of the assessed or assumed distribution type).

Both [1] and [2] require special treatment of each u_i . For example, [1] “dialing out” a bias might mean we have added just that – a “dial” or “free parameter”, and we have to count that. On the other hand, if we want to use u_i at say “2 sigma” instead of “1 sigma” coverage (95% vs. 68% confidence for the normal distribution) we might want to multiply s_i by two but leave b_i unaltered. So we have to separate these terms back out b_i , multiply s_i by two, and then recombine. But then, it is subsequently still a good idea to multiply this new U_i by a small-sample correction factor as we will see below in Section 3.

In its simplest form, we must quantify the following four types of uncertainty terms during the Model V&V process:

u_o : The uncertainty in the experimental data and measurement of the output QOI^{5,49}.

u_N : The numerical uncertainty in the spatial, temporal, and iterative solution domains, obtained from Solution Verification^{13,30,31,43}.

u_P : The parametric uncertainty in the model solution output, due to known or estimated uncertainties in the input parameters (geometry, material properties, boundary and loading conditions, etc.) This is usually obtained via sensitivity coefficient analysis or Monte Carlo style sampling methods to get *model* sensitivity to each input and their combinations, multiplied by input distributions to get u_P . In the process of obtaining u_P we also get Importance Factors^{6,7,25,27,28,49}.

u_M : The model form uncertainty. This elusive⁷ term contains all the remaining uncertainty not accounted for or underestimated in the above terms. Model form uncertainty may be due to missing physics, simplistic approximations, underestimating the width of the distribution of one of the other uncertainty terms, etc. It may be possible to assess u_M by Integral^{28,37} or Top Down Validation⁴⁰, or by Cross-Validation²¹ or assessment of Predictive Capability^{31,41,49}.

It seems legitimate by definition to view the total standard uncertainty as some combination of these four terms, since the last term, u_M , is a subjective “catch all” term called “model form”. The meaning and quantification of u_M in particular continue to challenge and frustrate the community, but some means of estimating u_M are at least possible if not universally accepted. A generalized combination of these four terms into a total standard uncertainty is given by:

$$\sigma_T = \sigma_T (\sigma_D, \sigma_N, \sigma_P, \sigma_M) \quad [8a]$$

In the *ideal case* where these four terms are separable and independent, we can express the total standard uncertainty¹² in our model’s characterization of the reality of interest (e.g. the product performing its job) as σ_T , where:

$$\sigma_T^2 = \sigma_D^2 + \sigma_N^2 + \sigma_P^2 + \sigma_M^2 \quad [8b]$$

Or, as used without correcting for small sample size, in the ASME V&V-20 nomenclature:

$$u_T = u_T (u_D, u_N, u_P, u_M) \quad [8c]$$

And again, in the *ideal case* where these four terms are separable and independent,

$$u_T^2 = u_D^2 + u_N^2 + u_P^2 + u_M^2 \quad [8d]$$

In reality, the situation is rarely this simplistic, and even when it is this simple and independent, the Model Validation procedure to assess values for the terms in Eqn. [8] is complex. However, the forthcoming ASME V&V-20 document provides methods and relationships for assessing these terms whether or not they are

independent or separable. These Standard Uncertainties may be obtained without making assumptions as to the form of their distribution (Normal, Uniform, etc.). It will be necessary to assess or assume forms for the distributions in order to make numerical assessments of confidence in reliability. For the moment, let’s assume the simple case, where the terms in Eqn. [8] are independent, and furthermore normally distributed, and develop the basic set of equations linking V&V with Reliability.

At this point, u_T is the term in the denominator of the Reliability Index β , so that

$$\beta = (M/u_T) \quad [3]$$

$$R = \Phi(\beta) \quad [4]$$

This treatment reduces conveniently to the case where, in a data-rich situation, the denominator contains only the product variability, that is, the term σ_D and an exactly correct assessment of the term σ_P . And if we assume for now that our model is exactly correct in its assessment of σ_P , and our assumption of normal distributions throughout is also correct, and if this is the one and only failure mode to be considered, we will obtain the correct Reliability of the product if the entire production run were to be tested in use.

It would appear that we are done. And, in the simplest case, where we neglect to assess the Confidence in our Reliability, and given the many other assumptions in this simplified situation, we are in fact done.

2.3.1 Assessing Confidence via Model V&V

Can we assume C=100% Confidence in our estimate of central margin $M = \mu_s - L$ and σ_T ? We would be a bit foolish to assume that at this point. If we consider the realm of confidence to be expressed in terms equivalent to the statistics of probability, then our Confidence could range from $0\% < C < 100\%$. Our reliability so far contains no assessment of confidence, so although it is our best estimate of R and a very precise estimate, so far, it is a deterministic value, a point estimate of reliability. Whether the true reliability is above or below this assessed estimate is a matter of a coin flip. To simplify, this is the same as saying we have a one-tailed Confidence $C_1=50\%$ (a coin toss) or equivalently a two-tailed Confidence $C_2=0\%$. That is, we estimate our reliability as a point value, but we know there is only an infinitesimal possibility it will take on that exact assessed value; even our model’s assessment only claims that the true reliability is either above or below our point estimate.

One of the Panel Discussion participants and later a reader of the draft of this narrative offered the following text that may be helpful in distinguishing Reliability, where we can assess that we approach C=100% Confidence, to the more common case where we know

we do not have C=100% Confidence, but we may not know a way, let alone a unique way, to quantify the confidence that we do have:

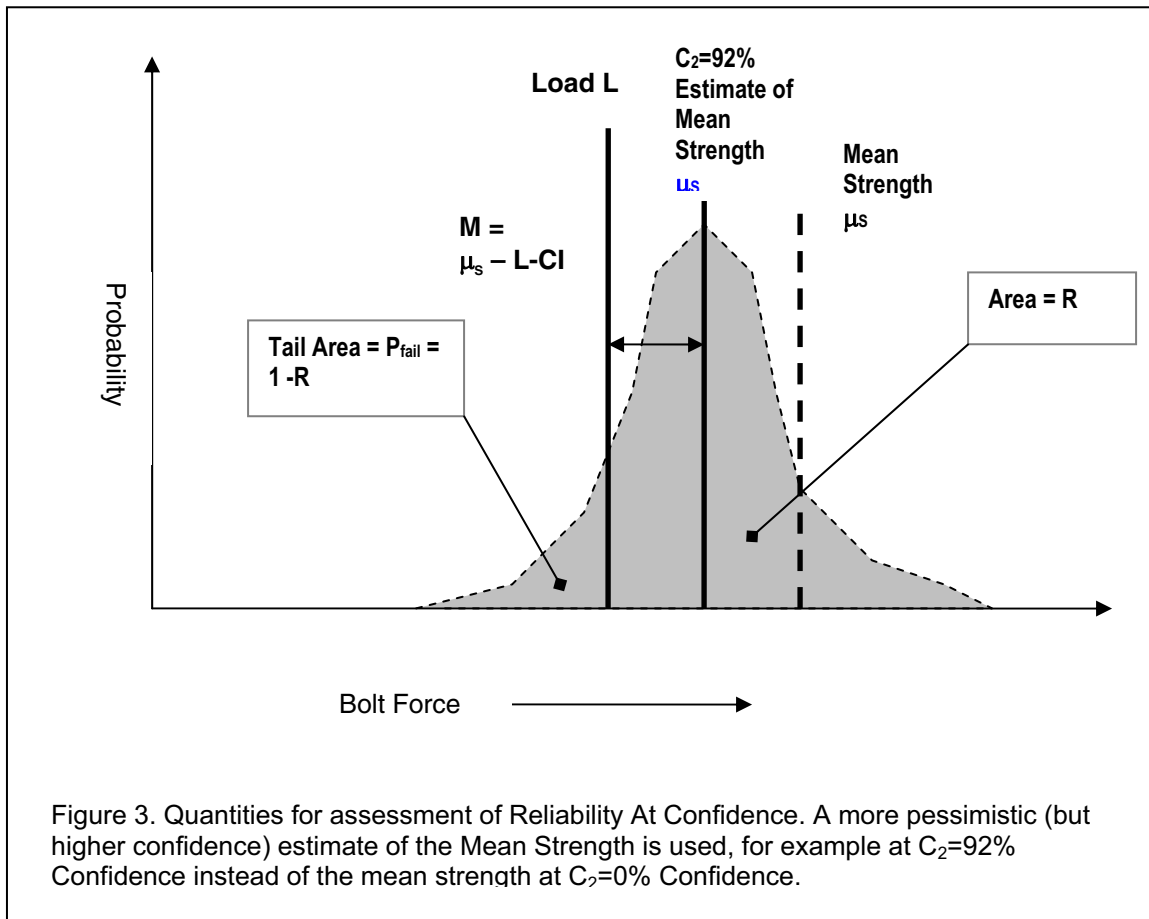
“Since reliability is a probability, the frequentistic (number of occurrences of an event in n independent experiments divided by n as n tends to infinity) and subjective interpretations of the probability of an event (a decision maker’s highest buying price of a lottery ticket that pays \$1 if the event occurs and zero otherwise) may be helpful to the reader. Many reliability studies rely on the subjective interpretation to construct models of uncertainty.”

In complex NDA, we will probably never escape subjectivity. However, we can hope that subjectivity becomes synonymous with expert judgment. Our goal in this portion of the narrative is to construct a simple quantitative roadmap from Model V&V to Reliability. Hopefully it is a way to quantify the process by which the decision maker chose his or her highest buying price for the “lottery ticket”. Although the assessment of confidence is often a *subjective* process, we feel it is important that assessment of confidence in reliability also be a *quantitative* albeit *subjective* process. Until we can find a way to quantify the confidence in our model assessed reliability, our procedure is missing a key component and remains open ended. Even our simplified procedure is going to take quite a few more steps to close this important loop from Model V&V to Reliability.

How do we assess our confidence in our estimate of central margin $M = \mu_s - L$ and $u_T \sim \sigma_T$, and what effect will this have on our point estimate of reliability? This is where the linkage between Model V&V and Reliability is essential. Without a validated model that quantifies its validation output at a specified confidence level, we cannot make this linkage.

First of all, we have to establish confidence in our estimate of each standard uncertainty term in eqn. [8d], $u_T^2 \approx u_D^2 + u_N^2 + u_P^2 + u_M^2$. What is our confidence in u_T ? Usually, we have not measured the population standard deviation, of any of the terms u_i , let alone their possible covariance. Instead, we have typically measured a sample standard deviation s_i . We want a confident estimate of σ_i instead, at whatever confidence we wish to assess the reliability. This is not the same as the typical use of a coverage factor, where $2s_i = 95\%$, $3s_i = 99.7\%$, etc. We still want an estimate of σ_i , the standard uncertainty (one standard deviation), but we want it at high confidence. Usually we do not have such high confidence because of a limited sample size. As a *beginning step*, Chi-Square correction factor may be used^{18,37,44} to recognize that the true value of the population standard deviation σ_i maybe greater than or less than the sample standard deviation s_i that was computed from the limited test data. In a non-probabilistic treatment, the computation of sample and population standard deviation may not be so simple, and methods such as Evidence Theory³³ or Generalized Information Theory³⁶ may be used.

Next, we have to examine our confidence in the numerator terms, μ_s and L . Let us assume for the moment that we know L exactly, as a product requirement, but that μ_s comes from our model, like some of the terms in Eqn. [8] also come from our model. Just like those terms, the confidence in μ_s will depend on validation of the model we used to obtain μ_s . Without model validation, we can take μ_s as the deterministic value of the mean capability, but we can only assume $C_1=50\%$ or $C_2=0\%$ Confidence. Even if the reliability from Eqn. [4] is assessed at $R=0.99$ or $R=0.9999$, the fact that we assessed that at $C_1=50\%$ Confidence is not reassuring. If we want higher confidence in our estimate of the Mean Strength μ_s , we will have to use a lower value for it, shifted downward by the width of a Confidence Interval (CI) in the mean, as shown in Fig. 3.



2.3.2 V&V Links Model Confidence to Assessed Reliability

The reliability at confidence is now assessed as before in Eqn. [3] and [4], except that we will correct the standard deviation σ_T by an epistemic (e.g. small sample) correction factor, denoted " X_c ". If we ignore the small sample correction, we would say that $X_c \equiv 1$. This is not the most optimistic assumption, since in reality, there is a chance that the true X_c is $X_c < 1$; we do not know the proper value of the small sample correction factor X_c , we can only start with a simple assumption and follow up with more elaborate analyses. One simple assumption is to let the correction factor $X_c = X_v^{-1}$, the inverse square root of the reduced Chi-Square term¹⁰. This gives σ_T at an assessed level of Confidence C:

$$\bullet \text{TIC} = X_c u_T. \quad [9]$$

The Margin M is also assessed at this level of Confidence C:

$$M|_C = \mu_s|_C - L = \mu_s - L - CI \quad [10a]$$

$$\text{FOSI}_C = \mu_s|_C / L = (\mu_s - CI) / L \quad [10b]$$

In the simplest of statistical assessments, we may obtain an estimate of the CI as^{24,31,32}:

$$CI = k t_c \bullet u_T \sqrt{\dots 0 + \frac{1}{N}} \quad [11]$$

In this case N =the number of experimental tests vs. simulations that are compared at the nominal values of input parameters (and therefore the nominal output) of the "Set Point"⁷ for Validation. (For the time being, we will only discuss the linkage of the Model and V&V to Reliability when the Model is used at this single "Set Point".) The quantity " k " is the (large sample) coverage factor⁷, intended to provide a CI at a desired Confidence=%C. For normally distributed estimates of the mean, $k=1$ for a 68% 2-sided CI, $k=1.96$ for a 95% 2-sided CI, etc. The quantity t_c is the *small-sample correction* factor for estimating the mean, since we may usually assume that the distribution of our estimate of the mean tends toward a normal distribution. Here, Student's $t = k t_c$. The first term under the radical is the number "0" [zero], to differentiate the CI Eqn. [11] from the PI Eqn. [15] below, which is identical except that a finite term will replace this "0". Again, for more complex situations, and for the multivariate case, more elaborate formulations^{11,38} are necessary, in a closed form