The Journal of the

SYSTEM RELIABILITY CENTER

Third Quarter - 2005

315.337.0900 General Information
888.722.8737 General Information
315.337.9932 Facsimile
315.337.9933 Technical Inquiries

src@alionscience.com via E-mail

http://src.alionscience.com Visit SRC on the Web

See page 22 for important changes in the RAC

# A Process to Accelerate Reliability Growth Testing that Provides Demonstrated Reliability With Statistical Confidence

*By: David C. Steiner, Timothy G. Ryan, Micah S. Koons, Raytheon Corporation, Space & Airborne Systems and Network Centric Systems, McKinney, Texas*

## Abstract

In this article, the authors present a process for accelerating reliability growth testing (RGT) that demonstrates an achieved level of reliability with statistical confidence. This process combines the best features of Highly Accelerated Life Testing (HALT) with the Army Materiel Systems Analysis Activity (AMSAA) Reliability Growth-Tracking Model [Reference 1]. Upper and lower operating limits for environmental stressors that are identified by HALT are used to run RGT under maximum allowed accelerated conditions. Times-to-failure (TTF) are then translated from accelerated time to normal-use-time or non-accelerated time with the use of scientifically accepted acceleration factors or functions. Elapsed test time is reduced while retaining a statistically significant demonstration of achieved reliability at confidence. The magnitude of the reduction in elapsed time depends on the failure mechanism having the lowest acceleration factor.

## Introduction

Proving that products meet a specified reliability requirement is getting more and more difficult. Reliability requirements, or the acceptable period between failures, keep getting longer, while the calendar time that is allocated to demonstrate those requirements keeps getting shorter. Today's development schedules often do not allow for a 6- to 12-month traditional growth or demonstration test. Keeping the status quo and insisting on long test times is problematic. For example, competitive pressures may require a product to be delivered before testing is complete, increasing the risk for both supplier and customer. Therefore, traditional demonstration or reliability growth techniques cannot continue to be used to show compliance to reliability requirements. To remain competitive in today's marketplace and minimize risk, the elapsed time it takes to demonstrate compliance to reliability requirements must be decreased. Yet, it must also be possible to do so with some degree of statistical confidence. The process presented herein accomplishes both of these objectives.

## Process

The process that has been developed to overcome these issues combines features of HALT testing and RGT. An overview of this process is shown in Figure 1.

The primary objective of the process is to run an effective RGT at the maximum operating limits of the hardware. HALT++ is used to identify the upper and lower operating limits of the hardware under test. These limits define the maximum stresses that can be placed on the hardware wherein the hardware continues to function to specification. This makes it possible to accelerate the test, collect all time-to-failure data and failure mechanism(s), translate failure times from accelerated time to non-accelerated time and to use these translated times as inputs to the AMSAA Model to make statistical statements about achieved reliability. Each step in the process now will be discussed.

1. **Define all system stress contributors.** The first step in the process is to clearly define all of the stressors that can act on the hardware in a lifetime of normal use service in the field. This

## INSIDE



System Reliability Center
201 Mill Street
Rome, NY 13440-6916

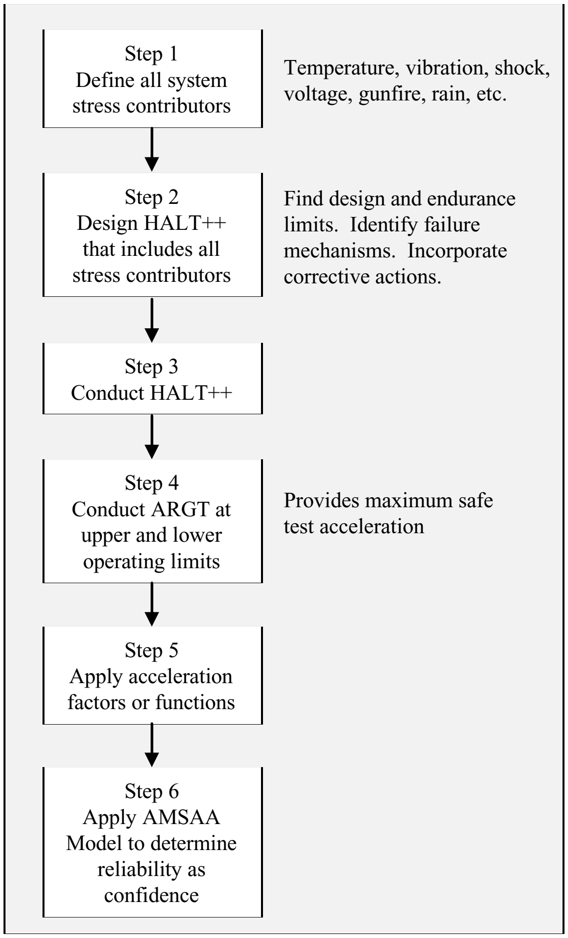



*Figure 1. Process Steps to Accelerate Reliability Growth Testing*

step is vitally important because low-test realism is often due to the omission of a relevant stress or incomplete definition of a type of stress. Avoiding this situation, and establishing a realistic test approach, requires a thorough knowledge of the life, mission, and environmental profiles that the hardware will experience over its lifetime. An effective test requires that the test conditions and procedures simulate these profiles. All environmental factors, natural or induced, which may affect system performance, must be identified and defined if the test is to be effective in eliminating weak links in the hardware or shortcomings in the manufacturing process.

It is only possible to obtain a thorough understanding of the stress contributors over a lifetime of service by reviewing all applicable program documentation and consulting with customer or end item users to identify and define all the operational, environmental, mechanical, and electrical conditions that are expected over the lifetime of the hardware. This task takes into account the range of the anticipated uses of the system and the range of the use environments for each of these conditions. In addition to the statement of work and equipment specifications, program documentation can include such items as mission planning, concepts of operation, and system requirements documents. The customer or user community can provide mission/life information, helping to assure that no stress contributor has been overlooked or over/under estimated.

Using the program documentation and customer and user inputs, the next action is to specifically analyze factors that impact long-term performance. For military systems, these factors include duration and timing of mission events for peacetime and wartime operating scenarios and conditions. It is important to note that conditions for each of these periods may be substantially different. We then evaluate all types of possible weather conditions and temperature ranges that could drive performance and operation, given the installation platform's local environments and levels. Other factors to be considered include possible power swings during operation, minimum and maximum voltage conditions, power cycling, ballistic shocks from gun firings, vibration conditions over all possible conditions, and other environmental factors that may be defined for possible locations and conditions where the system may be operated. Again, working closely with the customer and users will decrease the risk of overlooking or underestimating any stress contributors.

After all of the life, mission, and environmental stress contributors have been captured, this information is taken and consolidated into a cyclic profile that is representative of a "typical" system's life and which can be executed in a test chamber over a given period of time. The life, mission, and environmental stress contributors are compressed into an operational test profile to as closely as possible, given available test facilities, replicate what the hardware will be exposed to over one service lifetime. To have the highest level of test realism and effectiveness, the test profile should be representative of the stress contributors most likely to be encountered in the field at the anticipated levels, durations, and frequencies that most closely simulate life, mission, and environmental scenarios.

As a minimum, the operational profile should include temperature, vibration, shock, and power cycling that fully represents a service lifetime. If total lifetime durations cannot be simulated in a reasonable period of time, the profile should come as close as possible to maximize test realism. The test conditions and levels should be applied simultaneously, in so far as is possible, in a manner as closely as possible representative of field conditions.

To compress test schedules, higher than expected stress contributor levels must be used. Consult with the thermal and structural experts on the program to prevent unwanted destructive or catastrophic testing. For Accelerated RGT (ARGT), the upper and lower operating temperature limits and the upper operating vibration limit, quantified in HALT testing, are substituted for the "typical" life/mission/environmental temperature and vibration conditions.

2. **Design HALT++ that includes all stress contributors**. After all significant stress contributors have been determined and packaged in a cyclic operational profile, it is necessary to develop a HALT++ test that includes all of these effects. Typically the predominant stress contributors in most applications will be temperature and vibration. As shown in Figure 2, the traditional HALT test takes temperature and vibration into account. However, environmental conditions beyond temperature and vibration need to be evaluated for applicability and included in

HALT++. Figure 3 is an example of a HALT++ test that includes all contributing stress factors.

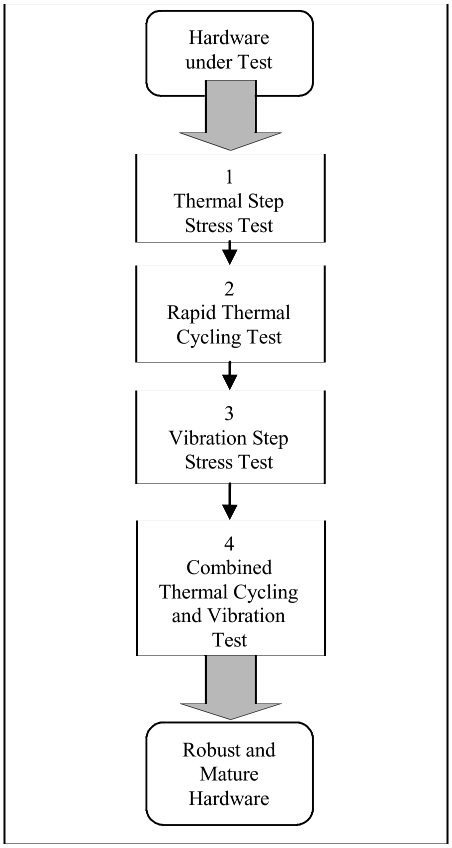



***Figure 2. Traditional HALT Test Sequence Testing***

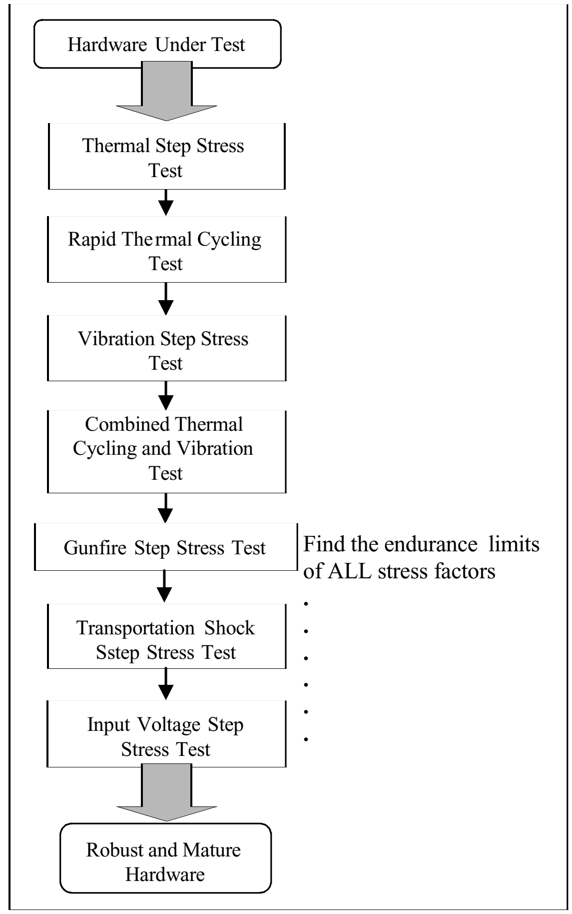



***Figure 3. Example of Advanced HALT++ Testing that Involves all Contributing Stress Factors***

It is important to work with personnel in the environmental testing laboratory to develop test regimens and procedures for testing needed conditions over and above those provided in the standard HALT chamber. For example, if humidity is an important environmental stress for the system under consideration, a humidity step stress test should be performed as part of HALT++ and the upper humidity-operating limit identified. If mechanical shock is a system stressor, increasing the frequency of the applied shocks can accelerate the shock stress. It is vital that all system stressors be included in the HALT++ test to identify maximum operating limits.

This HALT++ test has two objectives: 1) to be a development tool for improving the reliability of the hardware by finding weak design margins and correcting them early in the development cycle, and 2) to be a process for establishing the upper and lower operating limits of environmental stressors such as temperature and vibration. The upper and lower operating limits will be used in accelerated testing. Because of the dual objective of HALT++, the hardware used for testing should be as representative of the production version as possible, ideally the first or second assembly built.

The test hardware, test set-up and chamber are prepared for testing like in any other HALT test. The hardware must be instrumented to monitor environmental conditions and equipment status. Chamber preparation shall ensure maximum exposure of the test hardware to the environments.

As part of final preparation, it is important to coordinate with personnel in the test area and program mechanical engineering. These resources can provide test support, test guidance, and structural/thermal review. They also can provide information on chamber capability, test effectiveness and can be used as a resource to help prevent the possibility of any catastrophic event from occurring to the hardware due to over test.

3. **Conduct HALT++**. As stated earlier, HALT++ is used for early identification of inherent product design weaknesses and to determine the upper and lower operating limits of the environmental stressors for use in the ARGT. As shown in Figures 2 and 3, it is a process in which products are subjected to progressively higher stress levels to find weak links in the design and manufacturing processes. The process characterizes the test sample over the limits of the test and quantifies the product operating margin and design margin. These margins are above product specifications and require root cause failure analysis of all failures, even if they occur above specifications, with corrective

action decisions made for each failure. The environmental limits identified in HALT++ are used to verify operating and design margins and to establish test limits for ARGT to demonstrate reliability at statistical confidence.

This HALT++ test is performed in accordance with company guidelines and policy for HALT testing. The process, procedure, and intent of this test are the same as traditional HALT testing. At the completion of this test: 1) the reliability of the product will have been improved via the identification and correction of failures and 2) the upper and lower operating limits to all system stressors will have been identified. An appropriate safety margin derating of 5%-10% is applied to the operating limits used in Step 4 (ARGT) to help prevent unintended failures during ARGT.

4. **Conduct ARGT at upper and lower operating limits.** ARGT exposes equipment to environmental stresses above those expected during its actual service life. To reduce test time and compress testing, all stress levels must be increased above those anticipated in actual service. In ARGT, the derated upper and lower operating limits on all system stressors that were identified in process step 3 are substituted for what would traditionally be used, the actual conditions in the operational profile. These limits define the maximum stresses that can be placed on the unit(s) under test and still have them continue to function to specification. The implementation, purpose, intent, and objectives of the test remain the same as that for RGT or Reliability Demonstration Testing, the difference being the stress limits that are applied to the hardware.

Failure modes stimulated by these stresses are analyzed for root cause. Failure conditions are monitored and the TTF and failure mechanism, based on failure analysis, are documented. Corrective actions are implemented during the test to prevent their recurrence. As fixes are introduced into the equipment and failure modes are eliminated, the reliability of the equipment improves. As each failure occurs, we will record the TTF and the cumulative hours on all systems at the time of each failure. As each failure is analyzed to root cause, the underling failure mechanism will be determined and an industry accepted acceleration factor for that mechanism is determined.

5. **Apply acceleration factors or functions.** In this step the failure events that occurred in ARGT that are to be used for determining reliability are multiplied by the acceleration transform factor and/or function for each failure mechanism in order to convert the accelerated TTF to non-accelerated TTF. For each failure event, this conversion from accelerated TTF to non-accelerated TTF is required to calculate reliability using the AMSAA Model. Results from the AMSAA Model are valid only with non-accelerated TTF.

In other words, we will create a table of failure number and cumulative operating hours at the time of each failure as if no acceleration were applied to the testing. Figures 4 and 5 illustrates this, by showing the failures as they actually occurred during accelerated testing and as they would have happened if testing had been performed without acceleration.

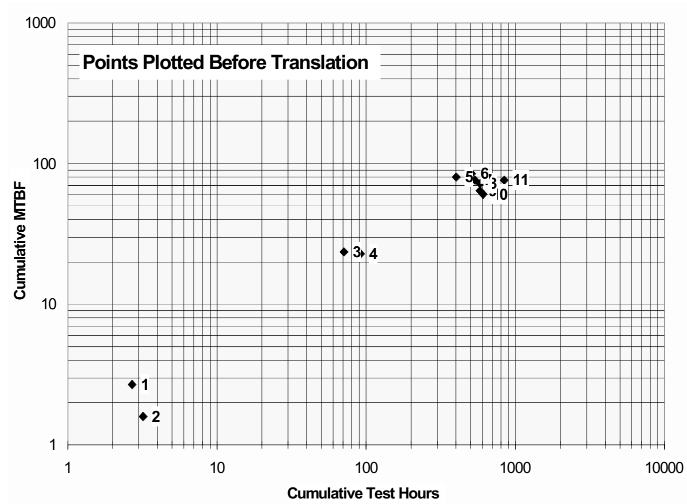



*Figure 4. Plot of Actual Failure Times DURING Accelerated Testing*

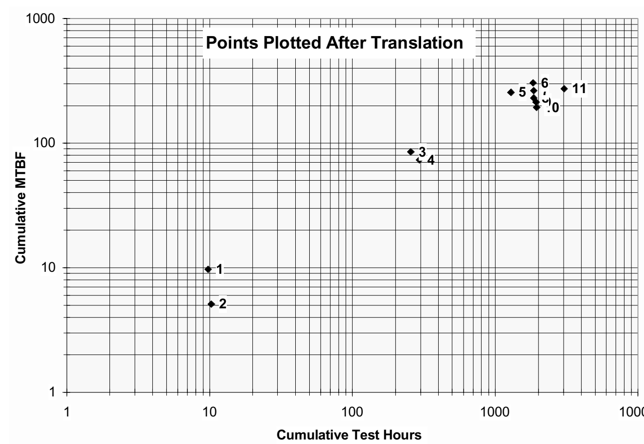



*Figure 5. Plot of Failures AFTER Translation*

Note: This is how the failures would have occurred if acceleration had not been applied. Note also that failures 7, 8, and 9 would have occurred in a different order.

This translation process is repeated during testing for each failure event and failure mechanism until all of the accelerated TTF have been converted to non-accelerated TTF. Each failure mechanism will have its own acceleration transform factor or function due to the physics of failure associated with each failure mechanism. Refer to Table 1 for an example of this conversion in an AMSAA-like input spreadsheet.

6. **Apply AMSAA Model to determine reliability at confidence**. Failures used to determine reliability that have been translated to non-accelerated TTF need to be input into the AMSAA Model to assess test status. These failures were converted to non-accelerated test times using the acceleration transform factor and/or function for the applicable failure mechanism in process step 5. This conversion was required in order to have data that is suitable for use in the AMSAA Model. After the failures have

been converted they are input into the AMSAA Model in chronological order. An example of an AMSAA Model data input spreadsheet populated with failure data is shown in Table 1.

**Table 1. Translation of Accelerated Time-to-Failure to Non-Accelerated Time-to-Failure Using Hypothetical Acceleration Factors**

| Failure Number | Time to Failure (accelerated hours) | Accel. Factor | Failure Accelerant | Time to Failure (real time) | Reordered Failure Number | Time to Failure (real time) |
|---|---|---|---|---|---|---|
| 1 | 2.7 | 3.6 | Temperature | 9.7 | 1 | 9.7 |
| 2 | 3.2 | 3.2 | Vibration | 10.2 | 2 | 10.2 |
| 3 | 70.8 | 3.6 | Temperature | 254.9 | 3 | 254.9 |
| 4 | 91.5 | 3.2 | Vibration | 292.8 | 4 | 292.8 |
| 5 | 400.9 | 3.2 | Vibration | 1282.9 | 5 | 1282.9 |
| 6 | 508.6 | 3.6 | Temperature | 1831.0 | 6 | 1831.0 |
| 7 | 534.1 | 3.6 | Temperature | 1922.8 | 8 | 1848.3 |
| 8 | 577.6 | 3.2 | Vibration | 1848.3 | 9 | 1848.3 |
| 9 | 577.6 | 3.2 | Vibration | 1848.3 | 7 | 1922.8 |
| 10 | 606.1 | 3.2 | Vibration | 1939.5 | 10 | 1939.5 |
| 11 | 838.5 | 3.6 | Temperature | 3018.6 | 11 | 3018.6 |

The vibration and temperature acceleration factors for Table 1 are as follows:

$$\text{Vibration Acceleration Factor} = \left[\frac{\text{Grms}_{\text{test}}}{\text{Grms}_{\text{normal}}}\right]^{4}$$

Source: [Reference 3, Exponent Reference 4]

$$\text{Temperature Acceleration Factor} = \left[\frac{\Delta\text{Test}}{\Delta\text{Tnormal}}\right]^{2.5}$$

Source: [Reference 3, Exponent Reference 4]

Translation of the TTF is straightforward. Given the accelerated time to failure and the physics of the failure mechanism, it is a simple process to translate each TTF to an un-accelerated test time, as shown in Figure 5 and Table 1. The acceleration factors shown in Table 1 are simple examples. In reality, each failure mechanism might be different for each failure. The only parameter required by the AMSAA model that is not simple to determine is the value for now (i.e., the current test time). In other words, given that the hardware under test has run for some period of time since the last failure, what value should be used for the current test time? There is no acceleration factor if there has not been a failure, so how can the accelerated non-failure time be translated to non-accelerated non-failure time? Several possibilities exist. What is recommended is using the average acceleration factor at the time of the last failure. By using this approach, there is a 50% probability that the acceleration factor is equal to or greater than average value. Using the example data shown in Table 1, after the 11th failure approximately 3.38 (i.e., [3.5*5 + 3.2*6]/11) hours are accumulated for each test hour until the test is completed or another failure occurs.

For the example test shown, testing stopped at 914 hours on success (no failure). Using the average acceleration factor of 3.38, this yields a stopping time of 3,090 hours in non-accelerated test time. We now have all of the parameters required by the AMSAA model. Reliability at confidence is computed using the AMSAA methodology with the applicable failure data input into the model in accordance with the requirements of the AMSAA Model. At the desired confidence level the model is run to calculate the reliability values provided by the model using the relevant test events. Relevant test events are those incidents that have occurred that are classified as failures to be scored for determining reliability of the hardware in accordance with program requirements i.e., mission-affecting failures, etc. The lower one-sided reliability value provided by the AMSAA Model is typically used to assess reliability status. This model also provides other statistical measures such as data accept/reject assessment, growth rate, instantaneous failure rate, lower/upper two-sided limit on MTBF, goodness of fit, level of significance, etc.

An example of an AMSAA Model output parameter data set is shown in Table 2.

**Table 2. Output of AMSAA NHPP Model**

| AMSAA Model Parameter | Value |
|---|---|
| *Beta ... growth rate or shape parameter* | 0.558 |
| *Lambda ... Scale parameter or intercept on log-log plot* | 0.124 |
| *Rho ... Failure Rate* | 0.002 |
| *Confidence on MTBF* | 0.8 |
| *Alpha (1-confidence)* | 0.2 |
| **Lower ONE-SIDED Limit on MTBF** | **337** |
| **Lower TWO-SIDED Limit on MTBF** | **285** |
| **MTBF ... Mean Time Between Failures** | **504** |
| **Upper TWO-SIDED Limit on MTBF** | **1,014** |
| *b ... unbiased estimate of the shape parameter* | 0.504 |
| *Goodness of Fit (GoF) Statistic* | 0.189 |
| *Chosen level of Significance* | 0.05 |
| *Critical Value (from Table 1)* | 0.214 |
| *Accept/Reject Model - GoF statistic < Critical Value* | **Accept** |

Steps 4, 5, and 6 are repeated until the unit(s) under test has demonstrated compliance to requirements or all members of the test team agree that testing can be terminated.

## Summary and Conclusions

We have presented a proposed process for accelerating reliability growth testing that demonstrates an achieved level of reliability with statistical confidence. The process combines the best features of HALT with the AMSAA Reliability Growth-Tracking Model. Upper and lower operating limits for environmental stressors that are identified by HALT are used to run a reliability growth test under maximum allowed accelerated conditions. We described a process by which times-to-failure are translated from accelerated to non-accelerated time with the use of industry accepted acceleration factors. Our example test data illustrates the proposed process, but we recognize that further process development is needed to enhance the accuracy and robustness of the process. The main areas for additional activity include development of scientifically accepted acceleration factors, accounting for uncertainty in both the acceleration factor(s) and the estimate for accumulating test time during failure free periods. Some of the key advantages and disadvantages of the process are listed in Table 3.

**Table 3. Advantages and Disadvantages of this Process**

| | |
|---|---|
| | **Advantages** |
| 1. | Statistical demonstration of reliability |
| 2. | Greatly decreased test time |
| 3. | Improved reliability |
| 4. | Incentive to identify problems and correct them as opposed to the reliability demonstration test which has rigid pass/fail criteria |
| | **Disadvantages** |
| 1. | Difficulty in identifying correct acceleration transform factors and/or functions |
| 2. | Accuracy of acceleration transform factors and functions |
| 3. | Open-ended completion date makes it difficult to plan as opposed to time terminated test |
| 4. | Accuracy of the estimate for accumulating test time during failure free periods |

## References


1. Broemm, William J., Paul M. Ellner, and John Woodworth, *AMSAA Reliability Growth Handbook, HDBK-A-1*, U.S. Army Materiel Systems Analysis Activity, September 1999.
2. Nelson, Wayne, *Accelerated Testing, Statistical Models, Test Plans, and Data Analysis*, John Wiley & Sons, 1990, pages 19 and 30.
3. Edson, Larry, *Demonstrating Reliability Requirements With Accelerated Testing*, 2004.
4. Blish, R.C., *Temperature Cycling and Thermal Shock Failure Rate Modeling*, IEEE, IRPS, 1997.


## About the Authors

Mr. Steiner is a Principle Engineer in the Product Integrity Engineering group. He provides reliability engineering expertise to programs across SAS and Network Centric Systems. He has 26 years of experience in all facets of reliability engineering. He has extensive hands-on experience in advanced development, full-scale development, and production programs. He has provided innovative solutions to numerous reliability problems. He is considered by his peers to be an expert in most reliability tasks and reliability mathematics and statistics. He holds a B.S. in Physics from Kansas State University.


**David C. Steiner**
Raytheon Space and Airborne Systems (SAS)
PO Box 801, Mail Station 8094
McKinney, TX 75070-0801


Mr. Koons is a Senior Systems Engineer in the Product Integrity Engineering group. Micah has a B.S. in Electrical Engineering from Texas A&M University and is an ASQ Certified Reliability Engineer. He has 26 years of reliability engineering experience and 7 years implementing MIL-STD-882 safety programs.


**Micah S. Koons**
Raytheon Space and Airborne Systems (SAS)
PO Box 801, Mail Station 8094
McKinney, TX 75070-0801


Mr. Ryan is a Senior Principle Reliability Engineer in the Network Centric Systems Department. Tim has a B.S. degree in Electrical Engineering from the Milwaukee School of Engineering and a MBA from the University of Dallas. He has over 28 years of experience in all facets of reliability engineering from physics of failure modeling to reliability testing. He has been responsible for the successful completion of numerous reliability programs on both radar and ground combat systems.


**Timothy G. Ryan**
Raytheon Network Centric Systems
2501 West University Drive, M/S 8094
McKinney, TX 75071-2813


# RMSQ Headlines

***Mission Possible … With Good Requirements***, Defense AT&L, published by the Defense Acquisition University, September-October 2005, page 20. In this article, the author, Wayne Turk discusses the "anatomy" of a good requirement, the requirements package, and potential traps in developing requirements.

***BAE Systems Talks Technology***, Aerospace Engineering, published by the Society of Automotive Engineers, August 2005, page 10. BAE Systems is exploring a number of technologies that will allow future pilots to control their aircraft without the use of conventional control surfaces. The benefits are increased stealth characteristics, reduced maintenance, higher reliability, and reduced weight. Control is achieved using airflow control. Ultimately, BAE Systems wants to use this and other technology breakthroughs to develop a maintenance-free aircraft.

***Testing toward Perfection***, Aerospace Engineering, published by the Society of Automotive Engineers, August 2005, page 31. Vought Aircraft Industries is using actual physical testing to verify and assess the results of design and analysis tools for airframe materials. The testing is called "a building block" approach and is used to stretch the current design envelope and demonstrate risk reduction at progressively larger scales while managing cost.

Starting with the expected environment, stress and test engineers identify candidate materials having the most efficient strength-

**Note: URLs and E-mail addresses in the Journal are hyperlinks. Click right on the hyperlink to visit a web site or send an E-mail.**

to-weight ratios. From these candidates, those materials having the other requisite characteristics (e.g., fatigue life) are selected for examination. Coupon-level tests are first run followed by progressively large scale tests on those materials that pass the previous level of testing. The tests are effective in helping understand how the materials fail. By identifying the modes of failure, design robustness can be improved.

(Continued on page 21)

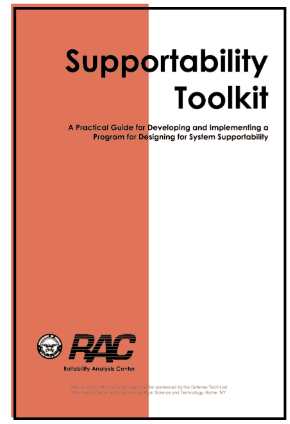
## Supportability Toolkit Now Available

Authored by two leaders in the areas of system engineering, logistics, and supportability, the Supportability Toolkit is now available from Alion. Dr. Benjamin Blanchard and John Langford developed the toolkit for Alion. The new toolkit is the fourth in a series developed by Alion and follows the Reliability Toolkit: Commercial Practices Edition, the Maintainability Toolkit, and the Quality Toolkit.

The Supportability Toolkit provides common sense and up-to-date information for planning and carrying out those activities needed to develop supportable systems and products and to support those products and systems in operation. It discusses the fundamental concept of supportability and related disciplines. It provides guidance on identifying and implementing the activities necessary to address supportability throughout the life cycle of an item. In today's environment, systems must be robust, reliable of high quality, supportable, cost effective, and responsive to the needs of the customer.

Click on this link to order the Supportability Toolkit on-line:

<http://src.alionscience.com/rac/jsp/webproducts/products.jsp>




## Independent Reliability Maturity Assessment - Improving the Bottom Line

Does your organization:

- Design or manufacture products for which high reliability is important?
- Warranty its products?
- Have a risk management program?
- Compete in a market where cost competition requires efficiency and effectiveness?

Then you should consider Alion's independent Reliability Maturity Assessment (RMA) services to:

- Identify your strengths and weaknesses for designing and manufacturing for reliability.
- Benefit from the lessons learned from a wide range of industries.
- Develop a roadmap for improvement.

An Alion RMA will help you improve profits and return on investment by:

- Producing products within budget and schedule to achieve reliability expected by customer.
- Controlling risk and warranty costs.
- Limiting liability.

### You Can Trust Alion's Experience and Expertise in Reliability

For over 37 years, Alion engineers and technical staff have supported a wide range of customers in both the public and private sectors by providing technical consulting and training in reliability. We have the experience of solving real-world problems, constrained by budget and schedule, for our clients. We combine the lessons learned from this experience with the latest techniques to solve tomorrow's challenges. Our RMA Approach is based on our experience, lessons learned, and up-to-date knowledge of the best practices for designing and manufacturing for reliability.

### Our Approach Provides Value-Added

Using our unique approach, Alion analyzes your company policies, procedures, and design and manufacturing processes, gains an insight into your company's culture of customers, competition, and product, and provides:

- A practical roadmap for improvement that includes short and long-term actions. You can start right away to quickly improve your process with small investments.
- A numerical rating that provides a snapshot of the maturity of a company's processes for designing and manufacturing for reliability.

Depending on the number and location of the company sites to be assessed, an Alion RMA usually takes 4 weeks, start to finish, with minimum interruption of company routine.

For information, please contact:

Ned H. Criscimagna
Tel: 301.918.1526
Fax: 301.731.0253
E-mail: <ncriscimagna@alionscience.com>

# DoD Assessments and Support

*By: Ned H. Criscimagna, Alion Science and Technology*

Alion Science and Technology supports one of the key missions of the Systems Engineering Directorate (SE), Office of the Under Secretary of Defense (Acquisition, Technology & Logistics)/Defense Systems. This mission is to conduct assessments of acquisition programs to ensure that systems engineering is an integral part of acquisition program execution.

The SE Assessments and Support (AS) office conducts assessments to improve the balance of cost, schedule, performance, and risk within and across programs that will operate in a system-of-systems environment.

The mission of the AS Office is to:

- Provide a context within which decisions can be made regarding individual programs.
- Help drive good systems engineering practices back into the way we do business.
- Assist program offices in implementing disciplined Systems Engineering, with integrated test and evaluation.
- Providing senior leadership with the comprehensive program assessments needed to support the decision making process.

The AS Office:

- Uses DoD staff resources possessing a wide range of experience and expertise from many organizations.
- Seeks to help program managers reduce risk through tailored application of an assessment methodology and development of specific recommendations.
- Conducts two major types of assessments: support and oversight.
  - Oversight assessments, provide independent, predictive views on the health of programs as part of the Defense Acquisition Board (DAB) process.
  - Support assessments are PM requested, with the resultant findings and recommendations developed exclusively for the PM's use.
- Ensures both types of assessments are constructive, providing actionable recommendations to position programs for success.

AS also conducts systemic analysis on the collected findings from multiple individual assessments. From this analysis, they develop a set of systems engineering best practices. These best practices are then shared with the acquisition community, which includes PMs, the military services' and OSD acquisition staffs, the Defense Acquisition University, industry, and professional associations.

A group of Program Support Team Leads (PSTLs) under the direction of the Deputy Director, Assessments and Support, manage the assessment efforts. The PSTLs are supported by personnel resources from government, industry, and academia. The organization of the AS Office is shown in Figure 1.

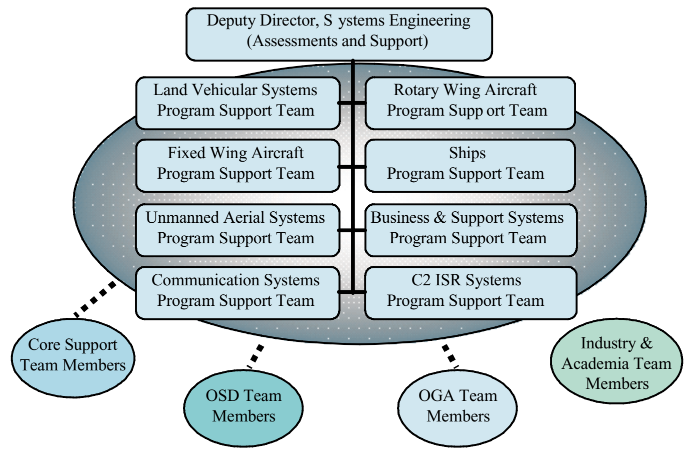



***Figure 1. AS Office Organization is Structured to Support Acquisition Programs by Type of System***

The PSTLs provide a single SE face to the PM and single SE voice to the DAB. They are supported by a diverse range of resources tailored to individual support requirements.

Alion engineers have been supporting the AS mission almost from its inception. Peter Tabbagh and Kenneth LaSala have been the principal engineers supporting the effort. Their focus has been on the reliability and maintainability aspects of systems engineering. They have directly supported a number of assessments including:

- F-22 Raptor
- F-35 Joint Strike Fighter
- V-22 Osprey
- Joint High Speed Vessel Program
- Expeditionary Fighting Vehicle
- Stryker Armored Combat Vehicle
- AH-64D Apache
- CVN-21 “21st Century Aircraft Carrier”
- UH-60M BLACK HAWK

Alion is proud to be part of the AS team contributing to the successful acquisition of critical weapon systems.

# Determining the Experimental Sample Size

*By: Jorge Romeu, Alion Science and Technology, Rome, NY*

## Introduction

A frequent question in experiments relates to calculating the required sample size "n." Engineers use samples to estimate or test performance measures (PM) such as reliability, MTTF, etc. Having an adequate sample size is important, for it determines the amount of time and dollars dedicated to the effort.

The sample size used in an experiment depends, first, on the statistical distribution of the random variable (r.v.) in question (e.g., device life). Such life may be distributed Normally, a symmetric distribution whose standard deviation is usually smaller than its mean and hence induces a moderate variability. Therefore, a smaller sample size can still yield an acceptable level of certainty (or uncertainty) regarding estimates and tests.

If the r.v. life is distributed say Exponentially, a highly skewed distribution having a standard deviation as large as its mean, the situation differs. Large variances induce large variability. Hence, the r.v. can now attain either very small or very large values. This fact introduces higher levels of uncertainty in estimations, which have to be compensated by drawing larger sample sizes. Therefore, inherent variability, or variance of the r.v. under study, constitutes the second factor of importance in sample size determination.

Finally, we have the issue of the level of "confidence" in estimation problems or of the Types I and II errors[1] in testing problems. To obtain higher confidence, all other factors being equal, we require wider "confidence intervals" (CI), which are usually not very useful. To reduce the width of a CI, we need to draw a larger sample. If instead of deriving a CI (estimation), we are testing, then we also need to consider the Type II error.

Summarizing, derivation of adequate sample sizes for testing or estimating a parameter requires three elements: the distribution of the r.v., its variability, and the risks of erring in the process of deriving such estimations or tests.

This article discusses and provides examples of several types of sample size derivations for location parameters. We first obtain sample sizes for interval estimation of the population using the Normal, Student t, and Exponential distributions, as well as for proportions. Then, we estimate the sample sizes for testing the mean of the Normal and of the Weibull distributions. Finally, we present examples of sample size derivation for the nonparametric (distribution-free) case. Due to its complexity, the derivation of sample sizes for estimating and testing variances is not addressed here.

## Sample Size for Interval Estimation of the Normal Mean

When a device life is distributed Normally, and we want to obtain an estimate of its MTTF, we base our sample size estimations on the formula for the CI for the mean ($\mu$):

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \bar{x} \pm H$$

In its half-length (H), which is the amount that is added and subtracted from the sample mean, $\bar{x}$, the CI formula includes four elements. The four elements are, the confidence level ($1 - \alpha$) desired, the random variation ($\sigma$) inherent in the Life of the device, the sample size (n) required to fulfill the requirements and the Normal Standard percentile ($z_{\alpha/2}$).

The preceding probability statement says that the CI will cover the true MTTF at least $100(1 - \alpha)\%$ of such times (e.g., 95% of the times). Let's assume we know the standard deviation of the population. Consider now pre-establishing a **fixed** CI half-length of $H = z_{\alpha/2}\times\sigma/\sqrt{n}$, about the true MTTF, for a pre-specified confidence level ($1 - \alpha$). Such equation H defines all our needs. After some algebraic manipulations, we obtain the sample size:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = H \Rightarrow n = \frac{(z_{\alpha/2})^2 \sigma^2}{H^2}$$

To illustrate this with a numerical example, assume that a device life is Normal distribution, and that the standard deviation $\sigma$ is known to be 8.6 units of time. Assume that we want to derive a 95% CI for the MTTF with a "precision" H of two time units (i.e., 95% of the times, MTTF estimates will be two units or less, away from the true but unknown MTTF of the device Life). Then, we would require a sample size of:

$$n = \frac{1.96^2 * 8.6^2}{2^2} = 71.03 \approx 72 \text{ observations}$$

Assume now that the device Life has a Normal distribution, but the standard deviation $\sigma$ is unknown, but estimated (s) from a pilot sample, prior experience, or using other means. Now we need to use an iterative process, using the Student t distribution:

$$\bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} = \bar{x} \pm H \Rightarrow n = \frac{(t_{\alpha/2,n-1})^2 s^2}{H^2}$$

[1]We commit a Type I Error when we decide that the alternative hypothesis ($H_1$) is true, when in fact the null ($H_0$) is true (e.g., assume that the mean is $\mu_1$, when in fact it is $\mu_0$). We commit a Type II Error when we decide that the null hypothesis ($H_0$) is true, when in fact the null ($H_0$) is true (e.g., assume that the mean is $\mu_0$, when in fact it is $\mu_1$).

The basic line of thought is exactly the same as before, except that now we use Student t instead of the Normal percentile. However, this introduces an interesting twist, since the t percentile requires knowledge of the Degrees of Freedom (DF), which in turn depends on the sample size (DF = n - 1). However, the sample size "n" is not known, because that is precisely what we are looking for with this procedure.

The solution is to set an initial, arbitrary sample size "n." Then, using the t percentile for these DF, we calculate a resulting sample size n'. Then, we compare n with n' and see if they agree or not. If they do, we stop. If not, we let n' determine the new DF and iterate.

We illustrate this method with the previous numerical example. But now assume that the standard deviation is unknown but we have an estimate of 8.6. Define an initial n = 20:

$$n' = \frac{\left(t_{0.05/2,20-1}\right)^2 8.6^2}{2^2} = \frac{2.093^2 \text{ x } 73.96}{4} = 80.998 \approx 81$$

Since n' = 81 differs from 20, we must iterate the calculations, using DF = 81 - 1 = 80:

$$n' = \frac{\left(t_{0.05/2,80-1}\right)^2 8.6^2}{2^2} = \frac{1.99^2 \text{ x } 73.96}{4} = 73.22 \approx 74$$

Proceeding this way, we arrive to the final value of n = 75, higher than the value n = 72, obtained for the case where $\sigma$ was known. Notice how we pay the price of drawing three additional observations to compensate for the lack of information about $\sigma$.

Finally, consider the case of Life with a Normal distribution, when the standard deviation $\sigma$ is known, or is unknown and estimated by "s." Suppose that, instead of deriving a CI, we require the sample size n for a hypothesis test. Then, in addition to Type I error $\alpha$, we must also consider Type II error $\beta$. Such error yields a difference of $\delta = \mu_0 - \mu_1$.

The sample size is obtained by considering a system of two equations, derived from the Operating Characteristic function (5), and assuming the two error probabilities $\alpha$ and $\beta$ are given. Solving the resulting system of two equations yields the required sample size:

$$n = \left[\frac{\left(z_\alpha + z_\beta\right)\sigma}{\delta}\right]^2$$

In the previous example, assume we now want the sample size n for a test that detects a difference of two units ($\delta = 2$) in MTTFs with errors $\alpha = 0.05$ and $\beta = 0.1$, when $\sigma = 8.6$:

$$n = \left[\frac{\left(z_{0.05} + z_{0.1}\right)\sigma}{\delta}\right]^2 = \left[\frac{\left(1.65 + 1.28\right)\text{x } 8.6}{2}\right]^2 = 158.7 \approx 158$$

By adding the extra requirement that we err if we accept a MTTF $\mu_1$ further than 2 units away from the true $\mu$, the sample size n has increased to 158. Error $\beta$ can now be at most 10% yielding $z_\beta = z_{0.1} = 1.28$. We have discussed these derivations in the START on OC Functions (5). The interested reader will find more details and numerical examples in that reference.

## Sample Size for Interval Estimation of a Proportion

A frequent query submitted to the staff at SRC deals with determining the sample size required for estimating the true proportion defective "p," or the true reliability "R" of a device, for a given Mission Time.

These two cases are, conceptually from a statistical point of view, handled the same. For, if we know the reliability required for a mission time "T," or we can estimate it, then a device failure to meet such reliability requirement is equivalent to it being "defective." Hence, the unreliability "p" is now P {Device Life < T} = p = 1 - R and, from the CI formula:

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}\left(1 - \bar{p}\right)}{n}} = \bar{p} \pm H$$

If, as before, we pre-establish the "precision" H, and the "confidence" (1 - $\alpha$) then, after some algebra, we obtain the formula for the sample size "n" required to fulfill these:

$$z_{\alpha/2}\sqrt{\frac{\bar{p}\left(1 - \bar{p}\right)}{n}} = H \Rightarrow n = z^2{}_{\alpha/2}\,\frac{\bar{p}\left(1 - \bar{p}\right)}{H^2}$$

To illustrate the percent defective (PD) calculations, assume we want to obtain the sample size required to estimate, with an 80% confidence ($\alpha = 0.2$) the PD in a production lot. Assume a precision "H" of, at most, 3% (e.g., the maximum distance we want such estimate "p" to be from the true, but unknown, lot PD, either by excess or by defect). We say: H = 0.03, with confidence (1 - $\alpha$) = 0.8. We need, as with the Student t case, a preliminary estimate of the true lot PD parameter. We can obtain such estimate from a pilot survey or historical data (or, worst case, by assuming p = 0.5). Let, in our example, this estimate be p = 0.05. Then, the sample size required is:

$$n = (z_{\alpha/2}/H)^2 \times p(1 - p) = [(1.28/0.03)^2]*0.05*0.95 = 86.47 \approx 87$$

Hence, a sample of n = 87 yields an estimate of PD "p", such that 80% of the time, it is not further than 3% from the true but unknown PD. The procedure is valid when n×p and n×(1 - p) are greater than 5. Our example is borderline, since n×p = 87 × 0.05 = 4.35 ≈ 5.

Assume we want instead an estimate of device reliability, for a Mission Time T, and we know it is somewhere around 0.95. If the reliability point estimate is: R = 1 - 0.5 = 0.95, then the probability of failure in time T is p = 0.05. Also assume a "precision" H = 0.03 (i.e., no further than 0.03 above or below 0.95) with at least 80% "confidence." Then, we perform the same calculations above (p = 1 - R) obtaining the same sample size n = 87.

An alternative method consists of using Binomial nomographs, which can be found in the References 1, 5, 10, 11, or 12. Nomographs are very useful in determining sample sizes when, if instead of a CI, we derive a hypothesis test. Then, in addition to Type I error $\alpha$, we must also consider Type II error $\beta$, which comes from accepting a bad hypothesis.

## Sample Size for Estimating the Exponential Mean

We know, from Reference (4), that if "n" devices have lives $X_i$, i = 1, … , n, distributed as Exponential with MTTF = $\mu$, then the statistic $2T/\mu$ (where $T = \Sigma X_i$) is distributed as Chi Square ($X^2$) with DF = 2n. From this we get the 100(1 - $\alpha$)% CI, say 95%, for MTTF ($\mu$):

$$\left( \frac{2T}{X^2_{2n,1-\alpha/2}} ; \frac{2T}{X^2_{2n;\alpha/2}} \right)$$

As in the Normal case, we want a "precision" or maximum distance "$\tau$" from either CI limit ($2T/X^2$) to the real (unknown) value of MTTF ($\mu$). But now, we express this as the ratio $\tau$:

$$\tau = \frac{\mu - \dfrac{2T}{X^2_{\alpha/2;2n}}}{\mu} ; \text{or; } \tau = \frac{\dfrac{2T}{X^2_{1-\alpha/2;2n}} - \mu}{\mu}$$

Following (1), we denote $C = X^2_{\alpha/2;2n}$ and $D = X^2_{(1-\alpha)/2;2n}$. We solve the preceding system of two equations for variables C, D, and $\tau$. After some algebra manipulations, we obtain:

$$\tau = \frac{C - D}{C + D} \Rightarrow \frac{C}{D} = \frac{X^2_{\alpha/2;2n}}{X^2_{1-\alpha/2;2n}} = \frac{1+\tau}{1-\tau}$$

Therefore, to obtain the adequate DF, we only need to inspect the Chi Square Tables, finding the ratio that fulfills the conditions, for confidence (1 - $\alpha$) and precision $\tau$. For example, assume we seek the sample size requirement for a 90% CI for the MTTF, with a precision of 45%. Then, 1 - $\alpha$ = 0.9, $\alpha$ = 0.1, $\alpha/2$ = 0.05, $\tau$ = 0.45 and ratio C/D is:

$$\frac{C}{D} = \frac{X^2_{0.95;24}}{X^2_{0.05;24}} = \frac{36.415}{13.848} = 2.62 \cong \frac{1+0.45}{1-0.45} = \frac{1.45}{0.55} = 2.636;$$

Hence: $2n = 24 \Rightarrow n = 12$

When the sample size n required is large, we can use the Normal approximation to the Chi Square distribution: $z = \sqrt{(2X^2_n)} - \sqrt{(2n - 1)}$. With some algebra, we then obtain:

$$\frac{\left(\sqrt{4n-1} + z_{\alpha/2}\right)^2}{\left(\sqrt{4n-1} - z_{\alpha/2}\right)^2} = \frac{1+r}{1-r} \Rightarrow n = \frac{1}{4} + \left(\frac{1}{2}\right) z^2_{\alpha/2} \left[ \frac{1}{\tau}\left( \frac{1}{\tau} + \sqrt{\frac{1}{\tau^2} - 1} \right) - \frac{1}{2} \right]$$

For example, assume we now seek the sample size requirement for a 90% CI for MTTF, with a precision of 20%. Then, 1 - $\alpha$ = 0.9, $\alpha$ = 0.1, $\alpha/2$ = 0.05, $\tau$ = 0.2. The result is:

$$n = \frac{1}{4} + \left(\frac{1}{2}\right) z^2_{0.05} \left[ \frac{1}{0.2}\left( \frac{1}{0.2} + \sqrt{\frac{1}{0.2^2} - 1} \right) - \frac{1}{2} \right]$$

$$= 0.25 + \frac{1.95^2}{2} \left( \frac{5 + \sqrt{25-1}}{0.2} \right) - 0.5 = 93.8$$

To verify this, we calculate the ratio of the two Chi Squares, with DF = 2n ≈ 188:

$$\frac{X^2_{0.95;188}}{X^2_{0.05;188}} = \frac{157.3}{221} = 1.41 \approx \frac{1+\tau}{1-\tau} = \frac{1+0.2}{1-0.2} = \frac{1.2}{0.8} = 1.5$$

Hence, a 90% CI for the Exponential MTTF of the device lives, with a precision of 45% of the true MTTF, i.e., $\tau$ = 0.45, would require drawing about 94 observations.

## Sample Size Requirements for Testing the Weibull Mean

Sometimes we need the sample size requirements for testing, instead of for estimating parameters. We will illustrate this situation for the Weibull distribution. Assume we need the sample size "n" to test the Weibull Mean Life "m", when shape parameter $\beta$ is known, and Types I and II errors (producer and consumer risks), device reliability R and test time T are given. Weibull also involves a scale or characteristic life parameter $\eta$ (now a "nuisance" parameter) that we, of necessity, need to substitute out of our equations.

We follow the algorithm described in (1). Using the Weibull density f(x), the cumulative distribution F(x), the mean life "m," and the Reliability R(x):

$$f(x) = \frac{\beta}{\eta^\beta} x^{\beta-1} \exp\left\{ -\left(\frac{x}{\eta}\right)^\beta \right\} \text{ and}$$

$$F(x) = 1 - \exp\left\{ -\left(\frac{x}{\eta}\right)^\beta \right\} \text{ and } m = \eta \times \Gamma\left(\frac{1}{\beta} + 1\right)$$

$$R(X) = P\{X \geq x\} = \text{Exp}\left\{-(x/\eta)^{\beta}\right\}$$

We can construct a Test Plan (n, c) that yields a sample size "n" and a critical number "c" (maximum failures to be observed) that fulfills the error and mission time problem requirements.

To achieve this, we assume that the r.v. "number of failures in test time T," denoted "x" can be approximated by a Binomial (n, p) distribution. The parameters are "n," the number of trials or devices placed on test, and "p," the probability of having a device failure at any trial:

$$p = F(T) = 1 - R(T) = 1 - \text{Exp}\left\{-(T/\eta)^{\beta}\right\}$$

We define a hypothesis test for device mean life "m", that fulfills Types I and II errors $\alpha$ and $\beta^*$, yielding Confidence (1 - $\alpha$) and Power (1 - $\beta^*$). The two hypotheses $H_i$: m = $m_i$ for i = 0,1 were originally based on the Weibull mean. However, they are now converted, after some algebra, into hypotheses $H'_i$: p = $p_i$ for i = 0,1, based on Binomial parameter "p":

$$p_i = 1 - \text{Exp}\left\{-\left(\frac{T}{m_i}\right)^{\beta} x \left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^{\beta}\right\}; i = 0,1$$

Since shape $\beta$ is known, reliability R(T) = 1 - p is only a function of T/m, the known test time "T" and the hypothesized Weibull mean "m." We can then establish a system of two Binomial equations that fulfill the required Types I and II errors (or risks) of the problem:

$$\sum_{x=0}^{c} C_x^n \, p_0^x \, (1-p_0)^{n-x} = 1-\alpha; \text{and}: \sum_{x=0}^{c} C_x^n \, p_1^x \, (1-p_1)^{n-x} = \beta^*$$

Solving this system of two equations, we obtain the appropriate values of "c" and "n" for the problem. This is the least preferred method, given its computational difficulties and trial-and error approach to obtaining simultaneously "n" and "c". We still use it (once "n" and "c" are obtained by one of the other two methods described below) but only to check their accuracy.

The alternative includes implementing a graphical method for obtaining such "n" and "c" values. It is similar to the method for obtaining an acceptance plan from an OC curve (5). Let's explain its use through a numerical example.

Say we seek the sample size "n" required to test that the mean "m" of a Weibull life is 5,000 hours, versus that is 1,000 hours. The time T available for testing each device is only 500 hours, and both risks $\alpha$ and $\beta^*$ are 0.01. The Weibull shape parameter is known to be $\beta$ = 2. We first need to calculate the two $p_i$, for i = 0, 1, from the equations given previously:

$$p_0 = 1 - \text{Exp}\left\{-\left(\frac{T}{m_0}\right)^{\beta} x \left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^{\beta}\right\}$$

$$= 1 - \text{Exp}\left\{-\left(\frac{500}{5000}\right)^{2} x \left(\Gamma\left(\frac{1}{2}+1\right)\right)^{2}\right\} = 1 - 0.9922$$

$$p_1 = 1 - \text{Exp}\left\{-\left(\frac{T}{m_1}\right)^{\beta} x \left(\Gamma\left(\frac{1}{\beta}+1\right)\right)^{\beta}\right\}$$

$$= 1 - \text{Exp}\left\{-\left(\frac{500}{1000}\right)^{2} x \left(\Gamma\left(\frac{1}{2}+1\right)\right)^{2}\right\} = 1 - 0.8217$$

Then, we place the two $p_i$ values obtained on the left scale of the Acceptance Plan graph (References 1, 5, and 7) in the Figure 1. Probabilities for Confidence (1 - $\alpha$) = 1 - 0.01 = 0.99 and Type II Error $\beta^*$ = 0.01 are placed on the right hand scale of the Acceptance Plan graph.

Finally, we draw the two connecting lines for these pairs of points, as done in Figure 1 (follows end of article), and find values n = 46 and c = 2, in the chart margins. These values were obtained by projecting the intersection point of these two lines, in the margin scales.

We can then check the resulting n and c values, by substituting them, jointly with the values $p_i$ for i = 0,1 and $\alpha$ and $\beta^*$, in the above Binomial equations. That is:

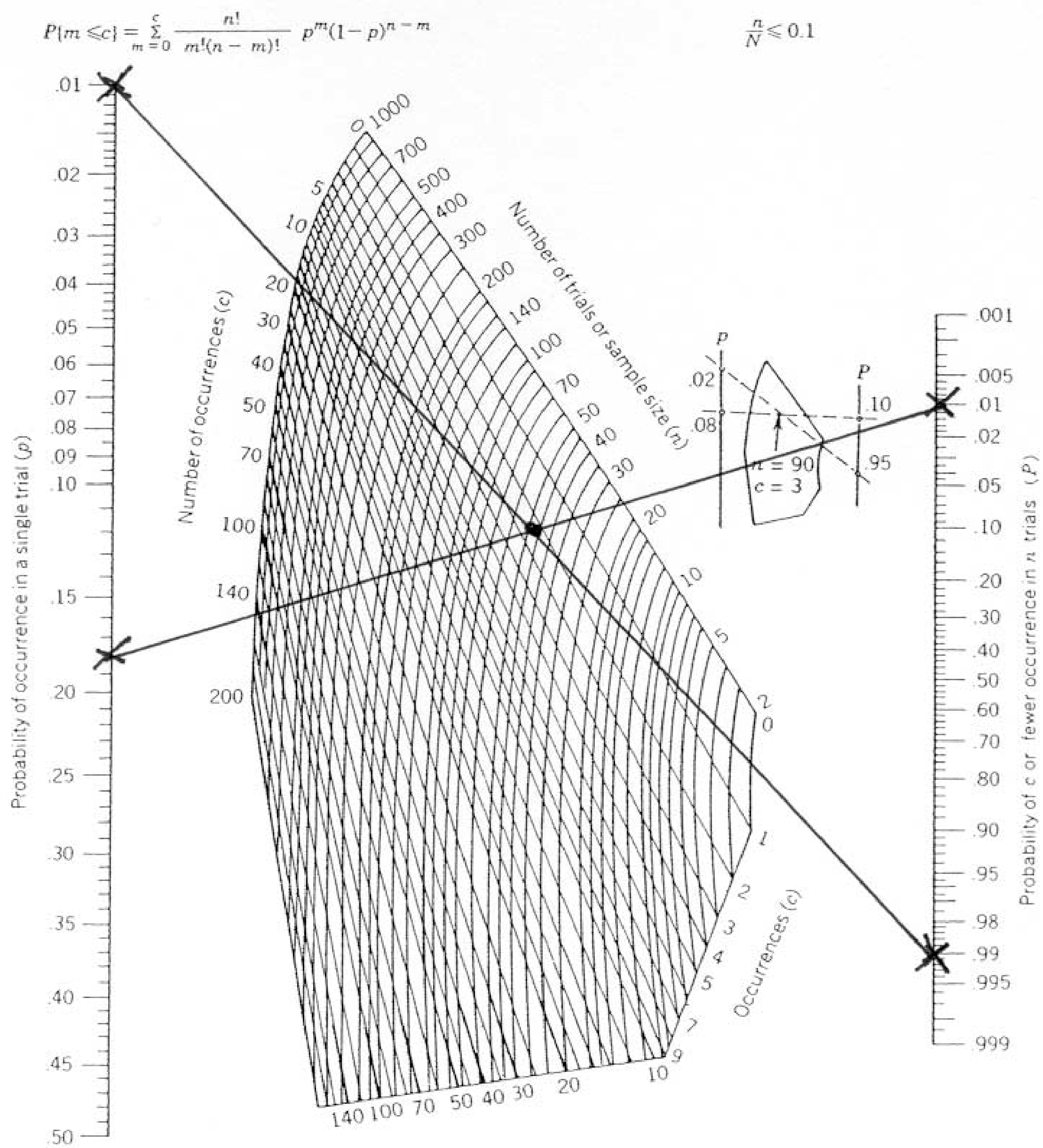
$$\sum_{x=0}^{c} C_x^n \, p_0^x \, (1-p_0)^{n-x} = 1-\alpha = 0.99;$$

$$\sum_{x=0}^{c} C_x^n \, p_1^x \, (1-p_1)^{n-x} = \beta^* = 0.01$$

There exists however, a third alternative or method for this problem, consisting in certain approximations that allow us to avoid the above graphical procedures. When sample size "n" is large, say greater than 20, the r.v. "x" approximates the Normal, with $\mu$ = np and $\sigma^2$ = np(1 - p). We can then, using the same two hypothesized $p_i$, for i = 0,1, and the two errors or risks $\alpha$ and $\beta^*$ given above, establish a system of two simultaneous equations to find adequate values for both n and c:

$$\frac{c - np_0}{\sqrt{np_0(1-p_0)}} = z_{\alpha};$$

$$\frac{c - np_1}{\sqrt{np_1(1-p_1)}} = -z_{\beta^*}$$




*Note:*
If $p$ is less than 0.01, set $k \times p$ on the $p$-scale and multiply the values on the $n$-scale by $k$, where $k = 0.01/p$ (taking $k$ to the next higher integer).

*Figure 1. Nomograph for Obtaining "n" and "c" Values*

(Continued on page 18)

## Determining the Experimental ... (Continued from page 15)

Here, the $z_\alpha$ are the Normal Standard percentiles for probability $\alpha$. Solving this system for "n" and "c", we obtain the equations that will yield the sample size and the critical number fulfilling the problem requirements:

$$n = \left[ \frac{z_\alpha \sqrt{p_0(1-p_0)} + Z_{\beta*} \sqrt{p_1(1-p_1)}}{p_1 - p_0} \right]^2$$

$$\text{and } c = np_0 + z_\alpha \sqrt{np_0(1-p_0)}$$

For the same numerical example given before, and substituting proportions $p_0 = 0.0078$ and $p_1 = 0.1783$ in the above equations, we obtain the adequate values n and c:

$$n = \left[ \frac{z_{0.01}\sqrt{p_0(1-p_0)} + z_{0.01}\sqrt{p_1(1-p_1)}}{p_1 - p_0} \right]^2$$

$$= \frac{z_{0.01}^2 \left(\sqrt{p_0(1-p_0)} + \sqrt{p_1(1-p_1)}\right)^2}{(p_1 - p_0)^2}$$

$$= \frac{2.326^2 \left(\sqrt{0.0078(1-0.0078)} + \sqrt{0.1783(1-0.1783)}\right)^2}{(0.1783 - 0.0078)^2}$$

$$= 41.3 \approx 42$$

$$c = np_0 + z_{0.01}\sqrt{np_0(1-p_0)}$$

$$= 42 \text{ x } 0.0078 + 2.326\sqrt{42 \text{ x } 0.0078 \text{ x } (1-0.0078)} = 1.66 \approx 2$$

We can verify how the three pairs of values (n, c), obtained by these three alternative methods, are very close, as expected.

## Sample Size and Nonparametric Estimation for Zero Failures

In the previous section, we discussed the case where we assumed that the device life is Weibull. Sometimes, we cannot (or do not want to) assume a specific distribution. In such cases, we must use a nonparametric approach (also known as distribution free, since no distribution is specified). However, there is a cost of not specifying a distribution: we now have to define the test length as equal to the Mission Time (not less, as we did above).

We again place "n" random, identically distributed, and independent items on a life test, now for the pre-specified Mission Time length "T." Each item will either fail or pass the test of length T. Hence, each item on test is an independent Bernoulli trial and the r.v. number of observed failures "x", out of "n" trials, is distributed Binomial. The failure probability is p = 1 - R, where R is the probability that any item survives mission time T.

Using the Binomial tables and the required reliability R, we calculate the sample size "n" that provides the "Confidence" (1 - $\alpha$) required in the problem statement.

For example, to demonstrate a reliability R = 0.95, with a Confidence 1 - $\alpha$ = 0.9, for a Mission Time of T hours and no failures, we place "n" devices on a test of length T. Each device can fail the test with probability p = 1 - R = 0.05. Since zero failures implies that all "n" devices "survive", we search the Binomial tables for a convenient sample size "n." This "n" must yield zero failures (c = 0), or equivalently twenty survivals, with Confidence 1 - $\alpha$ = 0.9. The Binomial (n, p) equation is then:

$$\text{Prob}\{\text{Obtaining "c" or less Failures}\} = \sum_{x=0}^{c} C_x^n p^x (1-p)^{n-x}$$

Since the required Confidence = 1 - $\alpha$ = 0.9 and zero failures is c = 0, we have:

$$P\{\text{no failures}\} = (1-p)^n = (0.95)^n$$
$$= (1-\alpha) = 0.9 => (0.95)^n = 0.0994 \text{ for } n \approx 45.$$

Hence, a sample of size n = 45 yields a confidence close to 0.95, of finding zero failures (c = 0) in a life test of Mission Time T, when the reliability for this mission time T is 0.95.

However, searching the Binomial (n,p) tables for a suitable "n" can be a tedious and time consuming task. We can instead use an equivalent equation, derived from such Binomial probability for the Confidence, for the case of zero failures (c = 0) or twenty successes:

$$\sum_{x=0}^{c} C_x^n p^x (1-p)^{n-x} = 1 - \text{Confence} = \alpha$$

$$\text{For } c = 0 \Rightarrow \sum_{x=0}^{c} C_x^n p^x (1-p)^{n-x} = (1-p)^n$$

$$= \text{Exp}\{n\text{Log}(1-p)\} = 1 - \text{Confidence} = \alpha$$

Taking Logarithms on both sides, noticing that p = 1 - R, and after some algebra, we obtain:

$$n = \frac{\text{Log}(1 - \text{Confidence})}{\text{Log}(1-p)} = \frac{\text{Log}(\alpha)}{\text{Log}(R)}$$

For example, applying this formula to the immediately preceding example, we obtain:

$$n = \frac{\text{Log}(0.1)}{\text{Log}(0.95)} = \frac{-1.0}{-0.2227} = 44.89 \approx 45$$

The results, obtained using the Binomial and the Logarithm formula, are close because both methods are totally equivalent. However, the second result (formula) is easier and faster to obtain than the first one (trial and error).

Summarizing, we first establish the problem requirements regarding the desired (1 - $\alpha$) confidence and acceptable reliability R. Then, we calculate the sample size n that satisfies these requirements. Such sample size n can then be used to estimate the reliability R, with the desired confidence. The life test must be of length equal to Mission Time T.

## Conclusions

The theory for determining the sample size that meets a testing or estimation requirement is extensive and complex. Such theory is driven by the type of parameter we want to estimate or test (i.e., location, scale, or shape) and by the distribution of the sampling statistic we use to implement the hypothesis test or to obtain the estimation.

In this article, we have discussed the problem of estimating and testing some location parameters (mean, proportion) for the Normal, Exponential, and Weibull distributions, and for distribution-free (nonparametric) situations. Our objective has been to illustrate the logic and the statistical thinking behind the derivation of such sample sizes. A better understanding of this logic may help practicing engineers to better implement such procedures.

We have only discussed a few of the most widely used cases. There are many other situations of interest. For a more extensive and in-depth treatment of this subject, the reader is referred to Chapter 13, pages 699 to 776, of Reference 1.

An assessment of the complexity of these derivations may be provided by the fact that the referred Chapter 13 is the last one of this extensive, two-volume reliability handbook. However, the manifold advantages that deriving an adequate sample size for our problem provides in terms of savings in time and effort, far outweighs its theoretical complexities.

## References


1. Reliability and Life Testing Handbook, Kececioglu, D., Volume 2, Prentice Hall, NJ, 1993.
2. Empirical Assessment of the Weibull Distribution, Romeu, J.L., RAC START, Volume 10, Number 3, <http://src.alionscience.com/pdf/WEIBULL.pdf>.
3. The Anderson-Darling Goodness of Fit Test, Romeu, J.L., RAC START, Volume 10, Number 5, <http://src.alionscience.com/pdf/A_DTest.pdf>.
4. Reliability Estimations for the Exponential Life, Romeu, J.L., RAC START, Volume 10, Number 7, <http://src.alionscience.com/pdf/R_EXP.pdf>.
5. OC Functions and Acceptance Sampling Plans, Romeu, J.L., RAC START, Volume 12, Number 1, <http://src.alionscience.com/pdf/OC_Curves.pdf>.
6. Quality Toolkit, Coppola, A., RAC, 2001.
7. Practical Statistical Tools for Reliability Engineers, Coppola, A., RAC, 2000.
8. Mechanical Applications in Reliability Engineering, Sadlon, R., RAC, 2000.
9. Statistical Analysis of Materials Data, Romeu, J.L. and C. Grethlein, AMPTIAC, 2000.
10. Probability and Statistics for Engineers and Scientists, Walpole, R., R. Myers, and S. Myers, Prentice Hall, NJ, 1998.
11. Introduction to Statistical Analysis (3rd Ed), Dixon, W.J. and F.J. Massey, McGraw-Hill, NY, 1969.
12. An Introduction to Reliability and Maintainability Engineering, Ebeling, C.E., Waveland Press, IL, 1997.


## About the Author


Dr. Jorge Luis Romeu has over thirty years of statistical and operations research experience in consulting, research, and teaching. He was a consultant for the petrochemical, construction, and agricultural industries. Dr. Romeu has also worked in statistical and simulation modeling and in data analysis of software and hardware reliability, software engineering and ecological problems.

Dr. Romeu has taught undergraduate and graduate statistics, operations research, and computer science in several American and foreign universities. He teaches short, intensive professional training courses. He is currently an Adjunct Professor of Statistics and Operations Research for Syracuse University and a Practicing Faculty of that school's Institute for Manufacturing Enterprises.

For his work in education and research and for his publications and presentations, Dr. Romeu has been elected Chartered Statistician Fellow of the Royal Statistical Society, Full Member of the Operations Research Society of America, and Fellow of the Institute of Statisticians. Romeu has received several international grants and awards, including a Fulbright Senior Lectureship and a Speaker Specialist Grant from the Department of State, in Mexico. He has extensive experience in international assignments in Spain and Latin America and is fluent in Spanish, English and French.

Romeu is a senior technical advisor for reliability and advanced information technology research with Alion Science and Technology previously IIT Research Institute (IITRI). Since joining Alion in 1998, Romeu has provided consulting for several statistical and operations research projects. He has written a State of the Art Report on Statistical Analysis of Materials Data, designed and taught a three-day intensive statistics course for practicing engineers, and written a series of articles on statistics and data analysis. He is Certified Reliability Engineer under the American Society for Quality certification program.

## RMSQ Headlines (Continued from page 7)

***Building a Balanced Scorecard***, Quality Digest, published by QCI International, September 2005, page 44. This article explores the ideas presented in The Balanced Scorecard by Robert S. Kaplan and David P. Norton (Harvard Business School Pres, 1996). The balanced scorecard is a model of metrics in four dimensions: financial, internal performance, customer and marketplace, and human resources. These categories of metrics drive performance across short, medium, and long-term time frames. The author, Craig Cochran, explains how the balanced scorecard approach is implemented and includes a short case study.

***The Myth of the Best Practices Silver Bullet***, CrossTalk, published by the Oklahoma City ALC, Ogden ALC, and the Warner-Robins ALC Software maintenance Groups, September-October 2005, page 14. The authors explore the key relationships that must be considered when applying best practices to projects.

***ISO 9000 in Service: The Good, the Bad, and the Ugly***, Quality Progress, published by the American Society for Quality, September 2005, page 42. The author discusses the challenges of getting ISO 9001 registration for a non-manufacturing service organization having electronic transaction processing as its main product. A major challenge of registration was the fact that the company, a Medicare claims processing firm, has nearly 3,500 employees at eight locations in 180 different departments.

***Six Ways to Benefit from Customer Complaints***, Quality Progress, published by the American Society for Quality, September 2005, page 49. The concept of listening to the "voice of the customer" is not new. However, applying what is heard to improve performance is not always done or not done effectively. In this article, the author provides six practical ideas that can help an organization improve business performance through effective handling of customer complaints.

***Advanced Materials for Manufacturability***, Aerospace Engineering, published by the Society of Automotive Engineers, September 2005, page 16. Two goals of the aviation industry are maintenance-free structures with long life and engines operating at higher temperatures for fuel-efficiency and performance. Both of these goals require advances in materials. At the same time, the manufacturability of components made of these advanced materials must be considered. This article discusses the increased use of titanium and other specialty metals as alternatives to the long-term standby, aluminum. The primary reason for this shift away from aluminum is the increasing use of graphite composites, which are galvanically dissimilar to aluminum.

***SAE100 Future Look***, Aerospace Engineering, published by the Society of Automotive Engineers, September 2005, page 42. In celebration of the SAE's Centennial, short discussions are provided the future of six technical areas. These areas are:

- Electric braking in military and commercial applications
- Non-Destructive Testing
- Electronic warfare
- Aerospace data communication
- Landing gear material
- Aerospace and avionics parts sourcing

These areas have the potential to improve reliability, safety, maintainability, network performance, costs, and supportability.

***Root Cause Failure Analysis***, Reliability, published by Reliability Magazine, LLC, Vol. 10, Issue 6, page 13. The author discusses six errors commonly made when conducting root cause failure analysis and provides ways of avoiding the errors.

# Open Training Announcement

Alion's next open registration courses are scheduled for February 7, 8, and 9, 2006 in Orlando, FL. The following courses will be offered:

- **Electronic Design Reliability** - taught by James Gormady.
- **System Software Reliability** - taught by Ann Marie Neufelder.
- **Reliability Engineering Statistics** - taught by Jorge L. Romeu.

The courses will be taught at the KDC-Orlando facility located at 8529 South Park Circle, Suite 150, Orlando, FL. The Orlando area is served by the Orlando International Airport (MCO). A number of hotels are conveniently located to the facility. A number of hotels are conveniently located to the facility.

For more information on the courses, please click on the following link: <http://src.alionscience.com/src/training.do>.

For more information on Alion's training program, please contact:

Ned H. Criscimagna<br>
Alion Science and Technology<br>
8100 Corporate Drive, Suite 400<br>
Lanham, MD 20785-2231<br>
Tel: 301.918.1526<br>
Fax: 301.731.0253<br>
E-mail: <ncriscimagna@alionscience.com>

# From the Editor

## Changes in the Former RAC

Welcome to the new electronic Journal devoted to reliability and related assurance disciplines. The change in the Journal is just one of many that we are making to the former RAC. Our new name, SRC, reflects the fact that we provide more than part-level reliability analyses and data. Certainly, the RAC was created to focus on part reliability data collection, analysis, and dissemination. For the past 37 years, however, we have expanded our capabilities and the services we offer to both commercial and government customers, and increased our emphasis on a systems perspective.

Today, the SRC is a full-service center of excellence that helps customers maintain high levels of system availability, reduce technical and schedule risks, and limit liability through applying system engineering principles and system integrity disciplines. The latter include reliability, maintainability, and supportability. We provide training, analytical, and consulting services in these disciplines.

We have 37 years of corporate experience, a staff of experienced engineers and technicians, and the data and tools to efficiently help our customers improve their bottom line and meet their customer and mission requirements. For government customers, for example the Department of Defense, improving the bottom line includes increasing mission reliability, system availability, and readiness while reducing life cycle costs. For commercial customers, improving the bottom line includes increasing profits and return on investment while reducing risk, liability, and costs.

We, of course, still have the most complete reliability databases available for components and systems. In addition, our new and copyrighted tool, SPIDR, System and Part Integrated Data Resource, provides a comprehensive, searchable database of up-to-date system and component field data. Together with our experience in conducting a wide range of projects for commercial and government customers, we use our data to as a tool in providing services that include:

- Performing and facilitating failure modes and effects and other reliability analyses.
- Conducting reliability maturity assessments.
- Developing and providing training.
- Designing effective tests.
- Conducting failure analysis.
- Developing useful tools, such as PRISM and SPIDR.

As for the Journal, the transition to an on-line version has been for me one of adapting to technology change. I am most comfortable dealing with hard copy. For me, hard copy has many obvious advantages. You can page back and forth to refer to a previous statement or figure, you can take and read it anywhere, and you mark up and highlight passages of particular interest.

*Ned H. Criscimagna*

However, hard copy also has many limitations. Distribution is slow and expensive. You cannot include active links that can, for example, take readers to the web site for a conference, advertisers' web sites, or sites with additional information on an article. You cannot keep the calendar up-to-date from month to month. Using multi-colors is expensive.

We have always limited the Journal to grays, whites, and reds. We no longer will be limited to those colors making graphics more interesting and easier to understand. We also are no longer limited to 24 pages, a limit that was based on economics. So we can include more articles and other features as they are available and can change the length from issue to issue to accommodate the material we have from contributors.

Yes, many changes are being made. But two things have not and will never change. The first is our dedication to providing our customers with the products and unbiased services needed to solve their problems, increase profitability, reduce costs, and meet company/mission needs.

The second constant is the quality and experience of our staff. We are proud of our low turnover rate, the experience of our staff, and the way that we continually hire new and young talents. We have some engineers, technicians, and support staff members who have been with us for more than 15 years; others have 25 or more years experience in reliability, maintainability, and supportability. Eight of our engineers are ASQ Certified Reliability Engineers and two are SOLE Certified Professional Logisticians. Our staff members belong to and participate in the Society of Automotive Engineers, The International Society of Logistics, the Institute of Electrical and Electronic Engineers, the International Electrotechnical Commission, and the American Society for Quality.

We are proud of our 37 years working in and contributing to the reliability and related assurance sciences. We also take great pride in the way we have helped our customers succeed, however an individual customer might measure success. We plan to build on our past, improving our capabilities to better serve you, our customers.

# Future Events in Reliability, Maintainability, and Supportability

**8th Annual Systems Engineering Conference**
October 24-27, 2005
San Diego, CA
E-mail: <pedmonson@ndia.org>
On the Web: <http://register.ndia.org/interview/register.ndia?PID=Brochure&SID_1GA0LCK>

**Aircraft Survivability 2005**
October 31 - November 3, 2005
Monterey, CA
Contact: Christy O'Hara
NDIA
Arlington, VA 22201
Tel: 703.247.2586
Fax: 703.522.1885
E-mail: <cohara@ndia.org>
On the Web: <http://register.ndia.org/interview/register.ndia?PID=Brochure&SID=_1HH0UP6>

**Military Nanotechnology**
October 31 - November 3, 2005
Contact: Jamison Nesbitt
The SMi Group
The Clove Building, Maguire Street
London, SE1 2NQ, United Kingdom
Tel: 44 (0)20 7827 6746
Fax: 44 (0)20 7827 6001
E-mail: <jnesbitt@smi-online.co.uk>
On the Web: <http://www.smi-online.co.uk/events/overview.asp?is=1&ref=2241>

**CMMI Technology Conference & User Group**
November 14-17, 2005
Denver, CO
E-mail: <pedmonson@ndia.org>
On the Web: <http://register.ndia.org/interview/register.ndia?PID=Brochure&SID=_1GH0WZW>

**2005 USAF Aircraft Structural Integrity Program (ASIP) Conference**
November 29 - December 1, 2005
Memphis, TN
Contact: J. Jennewine
Universal Technology Corporation
Dayton, OH 45432-2600
Tel: 937.426.2808
Fax: 937.426.8755
E-mail: <jjennewine@utcdaython.com>
On the Web: <http://www.asipcon.com>

**52nd Annual Reliability & Maintainability Symposium (RAMS)**
January 23-26, 2006
Newport Beach, CA
Contact: David F. Barber, Jr.
Tel: 828.898.6375
Fax: 828.898.6379
E-mail: <chair@rams.org>
On the Web: <http://www.rams.org/exhibits>

*For a complete listing of upcoming RMS events, click on the following link: <http://src.alionscience.com/newsandevents/src_calendar.html>*