

Establishing Effective ORT Requirements

Fred Schenkelberg, Ops A La Carte, LLC

Key Words: Ongoing Reliability Testing (ORT), Reliability Testing, Applied Reliability

SUMMARY & CONCLUSIONS

In some cases the use of reliability testing to sample the products at the end of assembly provides an effective means to detect shifts in materials and processes that adversely impact product reliability. Ongoing Reliability Test (ORT) design is a balance considering cost, timeliness, resolution, and accuracy. A poor ORT is costly and may increase the risk of significant field failure by falsely building management confidence. Or, the testing may be unable to detect even major adverse changes in the field failure rates by not evaluating the appropriate risks or with insufficient sampling.

This paper steps through the design of an effective product ORT program for a high volume consumer product. The analysis includes business objectives, design risk, vendor variability, and accelerated life testing considerations, while also considering the real factory constraints concerning equipment, skill, and time.

Considering the constraints and the major decisions based on the testing results permits the ORT to become an effective part of process control while providing protection from unwanted field failures. An effective ORT evaluates sufficient number of samples, with a set of stresses that accelerate the appropriate failure mechanisms and produces test results in a timely and informative manner

1 INTRODUCTION

The best post-production testing is the testing the does not have to occur. When the design, supply chain and manufacturing processes are robust, in control and the likelihood of a product incurring a field failure is sufficient low the benefit of on-going reliability testing is minimal. When the design has elements close to their apportioned reliability goals, or when the supply chain may provide highly variable materials, or when the manufacturing process is not in statically process control, then the use of ORT may provide significant benefit.

ORT provides a means to detect shifts in the materials, supplied components or assembly process that adversely affects the field reliability performance. The ORT is useful if it meets two criteria. First, it must cost less than the potential cost of the field failures averted. Second, it must effectively detect the desired defect rate threshold or the presence of an unwanted high defect rate within the population. Key aspects of the ORT design include effectiveness, timely, accurate and sensitive.

While working with a small team designing and producing a high volume consumer product, one of the stage gate criteria was the publication of the ORT plan for hand off to the manufacturing facility for execution. While this was standard practice for each product, there did not exist a set of criteria or instruction on how to create the ORT plan. This paper steps through the process of creating that ORT plan. It illustrates a real situation, including constraints, expectations and capabilities. While this process outline may not work in every situation, it is intended to further the practice for the creation of effective ORT requirements.

2 EXISTING & PROPOSED APPROACH

As the new reliability engineering within the group, the task of developing the ORT plan was naturally added to my list of tasks to accomplish prior to the release of the design to production. Naturally seeking guidance concerning the expectations and definitions of ORT within the organization, found two sets of conflicting information. One from the experience a test-engineering manager that had prepared a dozen or more ORT plans for previous projects. The second set of guidance from the new reliability engineering group manager with no experience developing ORT plans within this particular organization.

2.1 Existing ORT Practice

One of the final tasks for the test engineering manager was the creation of the ORT plan for execution by the manufacturing quality department. The test engineering team has worked closely with the design team to verify and validate every aspect of the products functionality and durability. And, as is common with a tight time-line product development schedule it seemed a Herculean task to accomplish all the final product verification and validation testing, documentation and reports, plus the myriad of details required for the handoff of the product to the overseas manufacturing facility. The ORT was often just one more task and not the most important when compared to finalizing the product for launch.

The common practice was to replicate portions of the verification testing most related to reliability. Drop testing, temperature & humidity exposure, thermal cycling, etc. and most were done to fixed standard based set of criteria. For example, the temperature & humidity exposure was set for 85°C and 85%RH for 168 hours. It was from a consumer product or military handbook standard and was routinely done

on every project. Failures never occurred, so it was considered a 'good' test.

It was expensive, it took a week to accomplish and the sample size/acceleration factor combination suggested the testing could at best detect a 50% failure rate with 60% confidence. There were no failures because the product was relatively robust or the sampling plan was not likely to detect a 10% failure rate after multiple rounds of testing. Many of the tests in previous ORT plans had similar issues. They didn't have sufficient samples, were not timely, failed to provide sufficient failure mechanism acceleration or didn't address areas of highest risk at all. Yet, the organization mandated that every product have an ORT plan executed during manufacturing.

2.2 Recommended ORT Practice

The reliability engineering manager recommend the ORT plan should be cost effective and able to detect unwanted high failure rates within the product prior to it's shipping to the customer. Continuing the discussion with colleagues, the design team project manager and the engineering manager, we developed the following approach to develop an effective ORT.

Define the product reliability goals and the apportionment of those goals to each major element of the product. And, define maximum acceptable defect levels.

Review the product shipping timeline from the time the product is ready for shipment till it reaches the final customer. For example, how long would the finished goods be within the manufacturing site?

Review product development testing results for process or supply chain variability, or for design elements with little margin. Prioritize on areas of highest possibility for field failures.

Review the manufacturing site's capabilities for performing ongoing reliability testing such as environmental chamber capacities. Estimate the cost per unit tested for each test.

2.3 Example information

To illustrate the ORT design approach let's use an example situation. A high volume consumer product that retails for \$50 has an expected production schedule of four million units over eight months. Using a linear production schedule (assumed for simplicity) that translates to production totals in a 3 shift, 7 days a week, factory to approximately:

6,000 units per shift

18,000 units per day

125,000 units per week

500,000 units per month

Furthermore, a unit moves along a life cycle starting with distribution channels till the end of the warranty period, as

follows:

Each day – shipment from factory to distribution centers (DC)

At 2 weeks – shipment from DC to retailers

At 1 month – sold from retailer to customer

At 1 year – end of warranty period

Of course the production and distribution assumptions above are simplifications. For the purpose of this paper we are ignoring production and distribution variations and complexities in order to focus on the process of designing an ORT that is cost effective. Adding the actual complexity of production and distribution would further refine the test plan.

3 DECISION CONSTRAINTS

There are a few key decisions related to the design of an ORT. These decisions impact the overall design of the test plan, plus shape the decisions that occur during testing. First, what is the risk we are willing to take shipping this product? Another way to phrase this question is; what defect level are we trying to detect with this test? This may or may not be the same as the associated reliability goal. The goal may be the target, yet due to time or budget constraints we are willing to invest in the ORT enough to detect a 'do not ship' threshold. Also, the acceptable risk may be variable and related to the cost of the failure. For example, a product failure that leads to liability lawsuits are much more expensive than a customer complaint or product return.

Assuming the product design is safe, we can focus the ORT design to detect an acceptably high field failure rate that would result in high warranty costs due to product returns. Before considering the effects of different product failure rates, consider that when a failure is caught in the lifecycle incurs different cost. After final assembly, the product accumulates additional costs that contribute to the cost of failure.

Assume a product failure caught at the factory simply incurs the cost of the materials and labor to produce the product. At each stage, the cost of failure increases, due to the additional costs for retesting, sorting, repackaging, etc. And, once with the customer and under warranty the costs include a replacement unit. For the purpose of this paper's example let's use the following cost of failure estimates associated with each stage (when the failures are identified):

Factory - \$25/unit

DC - \$30/unit

Retailer - \$50/unit

Customer - \$100/unit

Costs may include labor, shipping, packaging, testing, sorting, scrap, repairs and are dependent on the nature of the

failure. For planning purposes using estimates or historical cost information for similar products provides a starting point to determine the balance between the test time to detect a failure and the cost of a failure exposure.

Naturally, detecting failures earlier is less expensive to resolve. This is due to both the additional expenses for shipping, testing, and so on, yet is also includes the consideration of how many units are at risk. The longer the ORT takes to detect an unwanted high failure rate, the more units have been produced and in route to the customer. In the example outlined above, if the ORT takes a month to indicate an issue, there are approximately 500,000 units across the distribution channel from the factory pending today's shipment to the retail store shelves. These units could be quarantined, retested and defective units removed. Add an additional month to the time to detect an issue, and there would be an additional half million units in customer's hands.

The number of units exposed is a function of the ORT duration or time required to detect an unacceptable failure rate. Figure 1 shows the relationship between the test duration and number of units exposed.

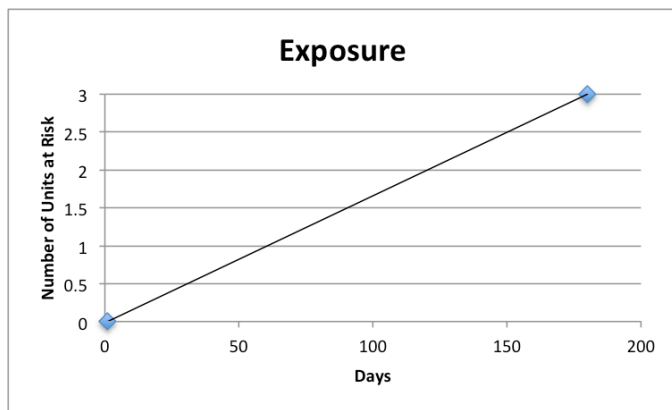


Figure 1 Plot of time versus number of units at risk

| Location | #Exposed | #failed | Failure Cost/unit | Cost |
|----------|----------|---------|-------------------|-----------|
| Factory | 18,000 | 360 | \$25 | \$9,000 |
| DC | 232,000 | 4,640 | \$30 | \$139,200 |
| Retail | 250,000 | 5,000 | \$50 | \$250,000 |
| | | | Total | \$398,200 |

Table 1 Cost of Failure

The cost of failures is related to the actual failure rate and

when the failure rate is detected. If the ORT detects a 2% failure rate in one day, the cost of failure would be 18,000 units (one day's production) times the failure rate time the \$25/unit (cost of failure while still in the factory), or, $18,000 \times 0.02 \times \$25 = \$9,000$. Whereas, if the same 2% failure rate was detected after one month of testing, the cost could be estimated as follows:

If it takes two months to detect the 2% failure rate, add to the total above the cost of the 2% failures of the 500,000 units in customer hands, for an additional cost of $500,000 \times 0.02 \times \$100 = \$1,000,000$ for a total of \$1,398,200. See Table 1 for results of all the calculations.

Compare this cost to the expected cost of failure generated by the acceptable 1% failure rate. The cost would only be incurred by customer returns, therefore the returns and cost would be 5,000 units at a cost per month of \$500,000. For a direct comparison recall that the units take a month to reach the customer, so the cost of a 1% failure rate over the same two months as the example above, would be \$500,000, compared to the \$1,398,200 for the 2% failure rate.

Figure 2 shows the relationship between test duration, cost of failure using select failure rates.

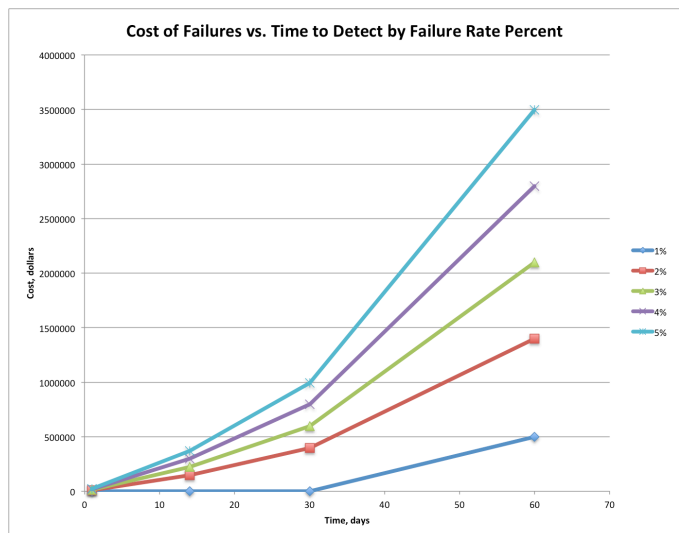


Figure 2 Cost of failures versus time

This brings up two related questions to resolve when building the ORT plan. What are the time constraints? And, what is the budget? The time constraint is in part a function of the ability to detect changes in failure rate. The next section using existing development testing results to build the ORT plan will include the detection consideration. The section on sample size directly relates to the testing budget.

4 DEVELOPMENT TESTING TO ORT PLAN

The development process includes the discovery of what elements of the design will limit the useful life of the product. From risk analysis to design verification, the design team learns about weaknesses in the product and production process including the supply chain.

The ORT plan takes shape as the product moves to production and we can use the information from past products, new product testing, and process and supply chain assessments. The idea is to select testing for the ORT that will illuminate product life issues that will occur over a period of time of interest. This period maybe the warranty period or the expected useful life of the product, in some cases, we may want to know the time to end of life. Whichever period is determined of interest, the selected testing should identify failures over that period of time. It is beyond the scope of this paper to discuss the range of accelerated life testing that may be employed for this testing plan.

With duration in mind, now review the history of similar products, if available, for the Pareto of field failures. The historical field returns reflect a similar product population's performance with actual customers and in their environment in real time, which is very difficult to replicate in the lab. They types of failures reflect the areas of possible weakness in the current design.

For example, let's consider a one-year warranty period Pareto from a similar product produced over the past two years. Figure 3 shows an example Pareto based on field returns.

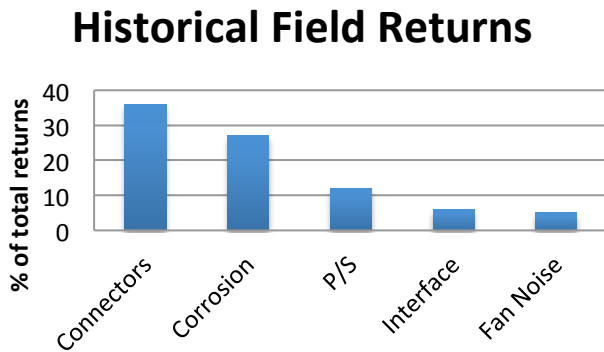


Figure 3 Historical Field Return information

In this example, let's assume the design team changed the type and robustness of the connectors among other design changes. And, product testing has shown significant improvement in avoiding bent pins thus reducing the risk for field failures. And, for this example, let's assume the new design addresses the corrosion issue with a new coating step in the process. Furthermore, the coating process window is small and even small variation in the process or materials results in an unacceptable coating, which would lead to premature product failure due to corrosion.

During product development the design verification testing for the new coating found that using a specific temperature, humidity and bias test (again beyond the scope of this paper to discuss the creation of a specific ALT) resulted in an acceleration factor of 12. In other words a 12-day test at accelerated conditions would replicate a year's worth of stress as seen in the field. Another piece of information, useful

during sample size determination, is the expected life distribution is a Weibull distribution with a shape parameter, β , equal to 3.

For the overall ORT plan, continue the analysis of failure information along with product design, supply chain, and process changes. Look for evidence that process or material variation would lead to a final product that is at risk for field failure. This provides the guide to selecting specific testing that will stress the product such that it will reveal product failures over the period of time of interest. It is this linkage between areas of greatest risk, the failure mechanism and the selected test that permits the ORT to detect the effect of assembled product variations on the product life.

5 SAMPLE SIZE DETERMINATION

The question of sample size includes consideration of the failure rate to detect, sampling confidence, test duration which is directly related to the acceleration factor above, and testing capacity. It is rare that an ORT will test as many samples for as long as desired to detect a small failure rate with high confidence. The practical limits of the cost of samples, testing facilities capacities, and the cost penalty of finding the unacceptable failure rate late.

A particular failure mechanism is only one way a system can fail. One way to use reliability apportionment is to consider the major failure mechanisms of a system and allocated a reliability target for each failure mechanism. This is like a reliability block diagram for system elements and major components, just along the lines of failure mechanisms. The results would be the same if each major element of a system has a single dominate failure mechanism.

5.1 Initial Calculations

For the purpose of the example, let's assume the reliability allocation for corrosion failures is 99% reliable over one year, or a 1% failure rate over the first year. Furthermore, let's assume the local policy concerning sampling a population is to use a confidence value of 90%. To get an estimate of the sample size, we can use the Success Run sample size formula [1] based on the binomial distribution.

$$n = \frac{\ln(1-C)}{\ln(R)} \quad (1)$$

Where,

n is the sample size

C is the statistical confidence

R is the reliability

Fixing confidence at three levels and plotting sample size versus reliability shows the increase in sample size needed for higher confidence and reliability. For a confidence of 90% to show at least 90% reliability, we need 22 samples. For higher reliability the sample size increases dramatically as is apparent in Figure 4.

Success Run Sample Size

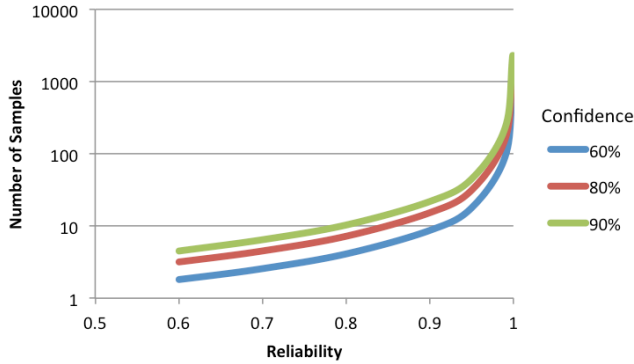


Figure 4 Sample size versus reliability by confidence level

Therefore, in order to detect a 1% defect rate that is introduced into the product, we would need to run 229 samples for two weeks under the determined test conditions to approximate one year of use.

5.2 Compromise Calculations

For the purpose of the example, let's assume the testing capacity is 50 units at any one time. Therefore, given 50 units, and solving equation 1 for reliability, we find we could detect a reliability of 95.5% over an approximate test time of two weeks.

Another option is to the samples longer to achieve a better resolution on reliability. Using a Weibayes formulation of the success run [2] to accommodate the Weibull distribution $\beta = 3$, the formula is equation 2.

$$n = \frac{\ln(1-C)}{m^\beta \ln(R)} \quad (2)$$

Where,

m is the number of lifetimes raised by β

β is the Weibull distribution shape parameter

and, the other terms are the same as in equation 1. Note: that when $m=1$ the equation reduces to that found in equation 1.

Running the test twice as long sets $m = 2$, with a $\beta = 3$, we find the 24 day test can detect a reliability of 99.4%. This is just over the desired goal of 99%, yet the testing takes twice as long, thus increasing the exposed number of units. Every 12 days add another 25 samples, such that every two weeks the month long results are available.

For this simple example, the tradeoff becomes detecting about a 4.5% defect rate in 12 days, or detecting a 1.2% defect rate in 24 days. The number of units exposed is 18,000 per day, or 216,000 and 432,000, respectively.

The cost of unreliability is slightly different since the longer duration test of 24 days will have some units exposed at

a higher cost per unit, plus, more total units exposed. One way to judge the two ORT plans is to assume the unknown true failure rate is 1%, 2.5% and 5%, then estimate the cost of unreliability based on the ability to detect the failure.

For the 1% failure rate, the test would confirm the acceptable low failure rate and no action is taken to isolate and inspect products. If the failure rate is 5%, both test plans would find the unacceptably high defect rate, and we can assume it would be just as the test ends, thus 12 and 24 days. For the 2.5% defect rate, let's assume the longer test takes the full 24 days to spot the higher than acceptable failure rate. And, let's assume the 12 day test would take at least two cycles to make the detection. To fully analyze the tradeoff, construct OC curves or run simulations may be useful. Of course the units under test may fail at any point during the test period, according to the probability governed by the associated Weibull distribution. If the failure occurs earlier than just at the end of the testing, we would receive an indication of an unacceptable failure rate and enjoy fewer units exposed. Another factor to consider is statistical error.

The acceptable field failure rate is 1%, thus any failures over that rate we can consider as additional cost that has gotten past the ORT. Therefore, for the cost calculations, we subtract the cost of the associated returns from the first 1%.

The following tables illustrate the comparison and associated costs assuming two failure rates, of 1.2% and 4.5% that are detected by the tests just as the test period ends.

| Actual Failure Rate | Cost of Unreliability (\$) | Marginal Cost of Unreliability (\$) |
|---------------------|----------------------------|-------------------------------------|
| 1% | 253,700 | |
| 1.2% | 304,440 | 50,740 |
| 4.5% | 1,141,650 | 887,950 |
| | Total | 938,690 |

Table 2 ORT plan with 24 day duration and 25 samples

Furthermore, let's assume it takes 4 2-week test periods to detect the 1.2% defect using the shorter test plan. Thus, there are 6 weeks of 1.2% defect rate product in the field producing 0.2% higher than acceptable returns at \$100 per return. This results in an additional cost for 6 weeks x 125,000 units per week x 0.002 failure rate x \$100 per return for an additional cost of \$150,000.

| Actual Failure Rate | Cost of Unreliability (\$) | Marginal Cost of Unreliability (\$) |
|-------------------------------------|----------------------------|-------------------------------------|
| 1% | 63,900 | |
| 1.2% | 76,680 | 12,780 |
| 6-weeks of 0.2% additional failures | | 150,000 |
| 4.5% | 287,550 | 223,650 |
| | Total | 386,430 |

Table 3 ORT plan with 12 day duration and 50 samples

The shorter test plan, while designed to catch 4.5% failure rates, reduces the total number of samples exposed to return over the target of 1%. The longer plan, in this simple

calculation, possibly costs nearly 3x as much in field returns, yet is much more capable to detect small changes in the failure rate.

6 CONCLUSIONS

The shorter test plan has fewer exposed units, which in this simple case has a total lower expected cost than the longer test. The longer test exposes more units to higher levels of the cost of unreliability, thus costs more. Of course, this is only based a very rudimentary calculation. Accounting for when the test actually has a failure and then calculating the exposure and associated costs would be a significant improvement.

In general, for ORT's evaluating an increase failure rate failure mechanism, the longer the test runs the better the ability to detect small changes in failure rate. Calculating the number of exposed units and the associated cost of unreliability for those units, provides a means to compare ORT plans. Often we are constrained by test capabilities, and using alternative test plans that fit within the constraints may increase the risk for higher than expected returns.

7 FURTHER WORK

Here are a few of ideas for further work to improve the process to design an effective and efficient ORT.

Simulate various perturbations in the underlying failure rate and determine the time for the various test plans to actually detect the change in failure rate.

Extend work to degradation ALT plans and to failure truncated plans.

Evaluate various failure mechanism patterns, such as wear out and infant mortality patterns.

Simulate the effect of various production schedules.

REFERENCES

1. Wasserman, Gary S. *Reliability Verification, Testing and Analysis in Engineering Design*. New York: Marcel Dekker, 2003, pg. 209.
2. Wasserman, Gary S. *Reliability Verification, Testing and Analysis in Engineering Design*. New York: Marcel Dekker, 2003, pg. 228.

BIOGRAPHIES

Fred Schenkelberg
Ops A La Carte, LLC
990 Richard Avenue, Suite 101
Santa Clara, CA 95050, USA
e-mail: fms@opsalacarte.com

Fred Schenkelberg is a reliability engineering and management consultant with Ops A La Carte, with areas of focus including reliability engineering management training and accelerated life testing. Previously, he co-founded and built the HP corporate reliability program, including consulting on a broad range of HP products. He is a lecturer with the University of Maryland teaching a graduate level course on reliability engineering management. He earned a Master of Science degree in statistics at Stanford University in 1996. He earned his bachelors degrees in Physics at the United State Military Academy in 1983. Fred is an active volunteer with the management committee of RAMS, currently the Chair of the American Society of Quality Reliability Division, active at the local level with the Society of Reliability Engineers and IEEE's Reliability Society, IEEE reliability standards development teams and recently joined the US delegation as a voting member of the IEC TAG 56 - Durability. He is a Senior Member of ASQ and IEEE. He is an ASQ Certified Quality and Reliability Engineer.