Planning Life Tests For Reliability Demonstration

by **William Q. Meeker, Gerald J. Hahn and Necip Doganaksoy**

How many units do I need to test and for how long to demonstrate high reliability?" Engineers and managers ask statisticians this question all the time because they need information about reliability before making important product design and product release decisions.

A reliability demonstration test shows—with a specified level of statistical confidence—the reliability meets a target value. It uses test results to determine a lower statistical confidence bound on reliability. If this bound equals or exceeds the target, demonstration is achieved.

Demonstrating high reliability for a complicated system is difficult with tests of reasonable size and length. System reliability models are, therefore, used to assess reliability. The reliabilities of the system's life limiting components provide important inputs to such models. These may be known from past performance or from physical considerations.

Often, however, our knowledge is limited, and empirical demonstration is required. As an example, we will look at a newly designed bearing for a washing machine, required to run flawlessly on 99% of all units for 4,000 cycles—corresponding to a conservatively estimated nominal usage rate of eight cycles (washes) per week for 10 years.

We will focus on "zero failure" tests—that is, reliability demonstration is achieved only if no units fail. Such tests are appealing because they require a minimal, but still often large, amount of testing.

If one or more failures occur, the desired reliability is not demonstrated, so such tests should not be started when it looks as though they won't succeed. Early testing needs to focus, instead, on speedily identifying and removing possible problems that compromise reliability.[1]

Making the Problem More Specific

The answer to "how many and how long" depends, in this particular application, on the reliability (R) we want to demonstrate over a specified

Use these formulas to determine sample size and test duration.

lifetime and the associated level of statistical confidence ($1-\alpha$). For this example, we need to demonstrate with $100(1-\alpha)\% = 90\%$ confidence that the reliability of the bearing is $R = 0.99$ at 4,000 cycles. In other words, the probability of failure by 4,000 cycles is to be no more than 0.01. For us to have a reasonable chance of achieving such demonstration, however, the actual reliability needs to be considerably larger than the level to be demonstrated.

Frequently, we also need to make an assumption about the time to failure distribution, but we will first present a test that does not require this.

A Demonstration Test With Minimal Assumptions

The simplest demonstration involves testing n units for the specified lifetime of 4,000 cycles. The demonstration succeeds if none of the n units fail.

This following simple formula will allow us to determine the required sample size:[2] $n = \log(\alpha)/\log(R)$. In the bearing example, this requires testing $n = \log(0.10)/\log(0.99) \approx 230$ units. Thus, if 230 randomly selected bearings run without failure for 4,000 cycles, we have demonstrated, with 90% confidence, a 4,000-cycle reliability of 0.99 for the sampled population.

The actual reliability needs to be approximately 0.999 to have an 80% chance of passing this test. That is, the probability none of the 230 units will fail is 0.999^{230}, or about 0.80.

A sample of 230 units, however, is prohibitively large in many applications.

A More Economical Plan

The number of test units can be reduced by running each unit beyond the specified lifetime—if we can make some assumptions about the distribution for time to failure based on knowledge of the failure mechanism and experience. The authors of *Statistical Methods for Reliability Data* consider the frequently encountered case when life follows a Weibull distribution with a known shape parameter β.[3] They show that, in this case, a zero failure demonstration test run for k multiples of the specified lifetime requires testing

$$n = \frac{1}{k^{\beta}} \times \frac{\log(\alpha)}{\log(R)} \text{ units.}$$

This is the minimal assumption plan when $k = 1$. If $k > 1$, the required sample size is reduced.

Application to Example

Experience suggests a Weibull distribution with $\beta = 2$ for bearing life. Under this assumption, the sample size can be reduced, relative to the minimal assumption plan, by a factor of k^2 by running the test for k multiples of 4,000 cycles. For example, taking $k = 3.4$ requires testing only 20 ($= 230/3.4^2$) units—but these all need to run failure free for 13,600 ($= 4{,}000 \times 3.4$) cycles to demonstrate 0.99 reliability with 90% confidence.

The actual reliability required to pass this test is the same as that for the corresponding minimal assumptions test. This follows from the relationship between the Weibull distribution with a known shape parameter and the exponential distribution.[4] Thus, to have even an 80% chance of passing the test, the actual reliability at 4,000 cycles again has to be 0.999.

Demonstration While Test Is Under Way

Figure 1 shows how the reliability that can be demonstrated increases as the test progresses. This chart shows, for an assumed Weibull distribution with $\beta = 2$ and $n = 20$ units, the 4,000-cycle reliability that is demonstrated with 50%, 80%, 90% and 95% confidence as a function of the number of failure free testing cycles up to 13,600 hours and beyond.

Figures 2 and 3 (p. 82) show curves to help plan a zero failure demonstration test for a component with a Weibull time to failure distribution. Figure 2 shows how n can be reduced by increasing test duration. It shows the required sample size as a function of k (multiples of lifetimes on test) for demonstrating a reliability of 0.99 with 90% confidence for $\beta = 0.8$, 1, 1.5, 2 and 3. In our example, $\beta = 2$ and $k = 3.4$, requiring $n = 20$. Figure 2 also shows increased values of β (implying less spread in the data) result in a smaller sample size.

Figure 3 shows the probability of successful demonstration as a function of the actual reliability for plans designed to demonstrate a reliability of 0.99 with 50%, 80%, 90% and 95% confidence. This probability does not depend on k, n or β, but requires testing a sufficient number of units for a sufficient time without failure to provide the required demonstration, such as $n = 20$ and $k = 3.4$ for $\beta = 2$. Figure 3 shows that if the actual reliability is 0.999, there is an 80% chance of successfully demonstrating a reliability of 0.99 with 90% confidence.

Assumptions and Pitfalls

Here are some further considerations in the planning and analysis of reliability demonstration tests.

Applicability of use-rate acceleration. With 75 cycles run daily, the test required six months. Such use-rate acceleration is possible for products such as washing machines that operate only a small fraction of the time. Also, usage cycle, rather than product age, was assumed as the appropriate scale for time to failure. This assumption often does not hold, even in simple situations. For example, time to failure for a photocopier depends on the number of start-ups and the number of copies per start-up.

Assumption of known Weibull parameter β. The assumption that $\beta = 2$ was critical to our example. When $\beta > 1$, the hazard function increases over time, suggesting product wearout.[5] Early failure modes, resulting in so-called infant mortality failures, are also assumed to have been eliminated. It is unlikely, for example, that our test could detect a manufacturing defect that causes early failure of 1% of the population.

In practice, β is rarely known. Using a value for β smaller than the actual value is conservative in planning a test (see Figure 2). Thus, if failure is due to wearout, using $\beta = 1$ leads to a larger sample than necessary and results in a higher actual confidence level than the nominal value.

Selection of test units. Careful attention needs to be given to the selection of the test units. These should represent the actual population of interest as closely as possible. Statistical theory calls for a random sample, but this may be hard to achieve when the available test units are specially built prototype units that may not be susceptible to some failure modes encountered in high volume production.

A poor practice. Some unexpected failure modes often do occur, and their root causes are usually addressed immediately to eliminate the failure mode in future production. It is then

FIGURE 2 Required Sample Size

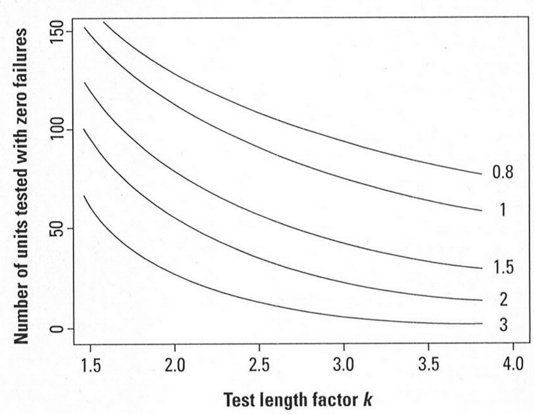
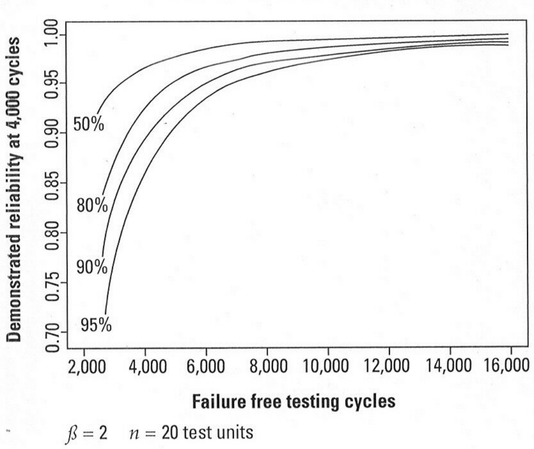
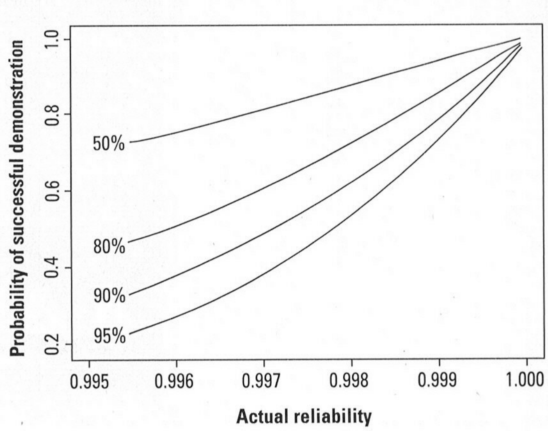


FIGURE 1 Demonstrated Reliability At Various Confidence Levels For a Weibull Distribution



$\beta = 2$ $n = 20$ test units

FIGURE 3 Probability of Successful Demonstration



tempting to proceed as if there were no failures, but this practice may be statistically biased and is wrong. The design fix may even introduce new failure modes. Redesigned units should always be tested.

Some extensions. Some extensions of the approach described here include:

- Tests that allow for failures. These require larger samples, but improve the probability of demonstration.
- Time to failure distributions other than the Weibull, such as the lognormal.
- Tests that assume no knowledge of the Weibull shape parameter.
- Procedures that use imperfect information about the Weibull shape parameter, such as Bayesian methods.

For more information on the first three topics, see *Statistical Methods for Reliability Data.*[6]

Other types of acceleration. Use-rate acceleration is not possible for a product that is in frequent use, such as a locomotive engine. In such cases, tests need to be conducted in more severe environments or with more severe stresses than encountered in actual use, such as accelerated temperature or voltage.[7,8] This calls for entirely different ways of planning a demonstration test—a topic we'll save for a future article.

REFERENCES

1. Harry McLean, *HALT, HASS and HASA Explained: Accelerated Reliability Techniques*, ASQ Quality Press, 2000.
2. G.J. Hahn, "Minimum Size Sampling Plans," *Journal of Quality Technology*, Vol. 6, No. 3, pp. 121-127.
3. W.Q. Meeker and L.A. Escobar, *Statistical Methods for Reliability Data*, John Wiley and Sons, 1998, sections 10.6 and 8.5.2.
4. Ibid.
5. Ibid.
6. Ibid.
7. G.J. Hahn, W.Q. Meeker and Necip Doganaksoy, "Speedier Reliability Analysis," *Quality Progress*, June 2003, pp. 58-63.
8. Wayne Nelson, *Accelerated Testing: Statistical Models, Test Plans and Data Analyses*, John Wiley and Sons, 1990.

WILLIAM Q. MEEKER *is a professor of statistics and distinguished professor of liberal arts and sciences at Iowa State University in Ames. He has a doctorate in administrative and engineering systems from Union College in Schenectady, NY. Meeker is a Fellow Member of the American Statistical Association and a Senior Member of ASQ.*

GERALD J. HAHN *is a retired manager of applied statistics at the GE Global Research Center in Schenectady, NY. He has a doctorate in statistics and operations research from Rensselaer Polytechnic Institute in Troy, NY, where he is also an adjunct faculty member. Hahn is a Fellow Member of the American Statistical Association and ASQ.*

NECIP DOGANAKSOY *is a statistician and a Six Sigma Master Black Belt at the GE Global Research Center in Schenectady, NY. He has a doctorate in administrative and engineering systems from Union College in Schenectady, NY. Doganaksoy is a Fellow Member of the American Statistical Association and a Senior Member of ASQ.*

Please comment

If you would like to comment on this article, please post your remarks on the *Quality Progress* Discussion Board at www.asq.org, or e-mail them to editor@asq.org.