

Taming the All-Equipment Reliability Test

Anthony Coppola, Senior Member IEEE

Key Words—Reliability testing, Production reliability, All-equipment reliability test

Reader Aids—

Purpose: Widen state of the art.

Special math needed: Elementary statistics.

Results useful to: Reliability engineers and program managers.

Summary & Conclusions—As currently used, the all-equipment reliability test provides unequal risks favoring the producer for short test periods and favoring the consumer for long test periods. This paper describes the empirical derivation of a method for modifying the test to provide equal producer and consumer risks over a range of test lengths. With the formulas provided in the paper, the user can structure an all-equipment test with equal s -risks for discrimination ratios of 2.0:1, 1.5:1, or 3.0:1, as desired.

INTRODUCTION

A critical consideration in invoking reliability tests is the s -risks involved. Each test plan has a s -risk of rejecting 'good' equipment (the producer risk) and a s -risk of accepting 'bad' equipment (the consumer risk). For a reasonable application, these s -risks must be known and acceptable to both parties, for quantified values of 'good' and 'bad'. As used here, 'good' equipment is that having a MTBF equal to the specified MTBF (Θ_0) as defined in MIL-STD-781B, and 'bad' equipment is that having MTBF equal to the minimum acceptable value (Θ_1). In MIL-STD-781C, these definitions are changed, but Θ_1 is still the lower test MTBF and Θ_0 , the upper test MTBF. Most of the reliability test plans in MIL-STD-781B have well defined producer and consumer risks; in fact, they are designed to establish desired s -risks based on the ratio of Θ_0 to Θ_1 . A notable exception is the all-equipment reliability test, Test Plan XXIX of MIL-STD-781B and its improved (2-line) version reported in the literature [1-3], used in current military contracts [4] and now incorporated into MIL-STD-781C. The risks provided by the test are a function of its length and the test can be extremely easy to pass (low producer risk, high consumer risk) or unreasonably stringent (high producer risk, low consumer risk). This paper describes a way to modify the test to provide approximately equal producer and consumer risks over a range of test lengths.

DESCRIPTION OF THE TEST

Since the 1-line all-equipment test described in MIL-STD-781B is superseded by a 2-line version in MIL-STD-

781C and since the 2-line version is now in common use, we will consider only the 2-line version. The test plan requires the testing of all equipments in a production run for a stated short interval of test time per unit. Failures are plotted against time (see Fig. 1). Two lines are drawn on the chart; the upper line (the y -axis represents failures) is a reject line and the lower a boundary line. If the plot of failures vs. time penetrates the reject line, the equipment is considered unacceptable and acceptance halted until corrections are made (in principle anyway). If the plot reaches the boundary line, it continues along the boundary line until another failure raises it above the line. The reason for this is to assure that any drop in production quality will quickly result in a reject decision, since the plot of failures vs. time is never further from rejection than the distance from the boundary line to the reject line.

Though the equations for the two lines could be arbitrarily derived, the general method has been to use for the reject line the same equation as the reject line in one of the sequential test plans of MIL-STD-781. The boundary line is drawn parallel to the reject line a given distance below and could be (but is not necessarily) the same as the accept line in the sequential test plan.

DIFFICULTIES WITH THE TEST

Since the test is never more than a given number of failures away from rejection, a degradation in production

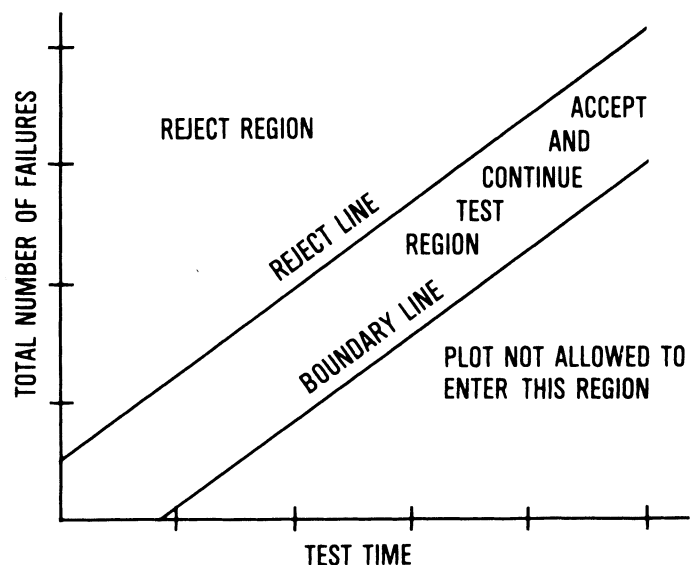


Fig. 1 All-equipment reliability test

quality can result in a quick rejection decision. However, if the test length is short, even poor equipment may not accrue enough failures to reach the reject line before the test ends. On the other hand, there is always a finite probability that satisfactory equipment will accrue sufficient failures (by statistical chance) to cover the distance between the boundary line and the reject line. This probability is negligible for any small amount of time, but as the test length grows, it becomes appreciable until the point is reached (for long tests) where even good equipment has little probability of passing. This ought to concern the consumer as well as the producer because rejecting good equipment is against the best interest of both parties. Example: An all-equipment test, using the formula $T = -3.46 + 1.386F$ to establish the reject line and the formula $T = 4.40 + 1.386F$ to establish the boundary line;

T test length in multiples of Θ_1
 F number of failures

These formulas are identical with those of the reject and accept line of Test Plan III of MIL-STD-781B, a sequential test providing 10% producer and consumer risks for a discrimination ratio of 2.0 ($\Theta_0/\Theta_1 = 2.0$). Equivalent forms are $F = 0.72T + 2.50$ for the reject line and $F = 0.72T - 3.17$ for the boundary line.

Figure 2 presents the operating characteristics of the test for total test lengths (number of equipments multiplied by time on each equipment) of 5, 10, 20, 40, 80 multiples of Θ_1 , and Table I converts this information to the producer and consumer risks.

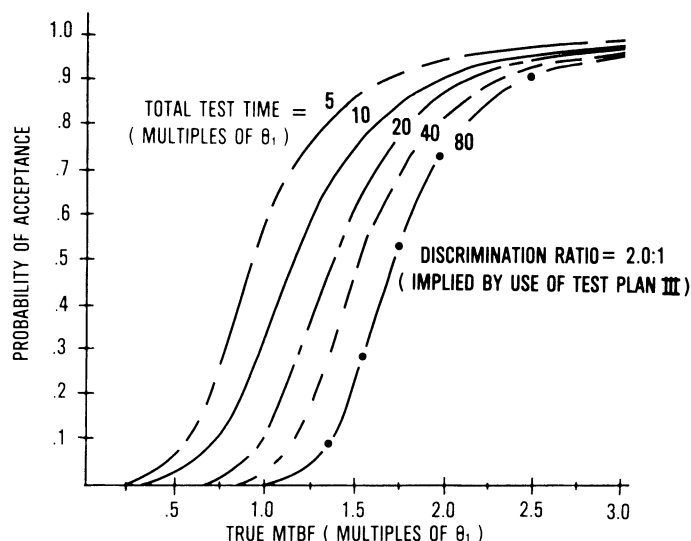


Fig. 2 Operating characteristics of the all-equipment reliability test based on test plan III

Table I

	Test Time (Multiples of Θ_1)				
	5	10	20	40	80
Producer Risk	.05	.09	.12	.19	.26
Consumer Risk	.57	.34	.12	.02	0

Notes:

Producer Risk = 1-Probability of accepting an equipment with MTBF = Θ_0 .

Consumer Risk = Probability of accepting an equipment with MTBF = Θ_1 .

As Table I indicates, the shorter test periods favor the producer and the longer the consumer (although, as stated above, the consumer ought not be pleased at high probabilities of rejecting good equipment).

The test is exceptionally poor in the shorter lengths. While short tests will always have greater s-risks than longer tests, the producer and consumer ought at least accept equal risks. Hence, a modification to the test plan equalizing the s-risks is warranted.

DERIVATION OF THE NEW TEST

The recommended new test plan is based on the hypothesis that equal s-risks could be obtained by varying the distance between the reject and boundary lines as a function of test length. This would be implemented by replacing the constant in each equation with a function of test length.

The procedures for deriving the function were strictly empirical.

1. The equations for a sequential test with nominal 10% risks and a 2:1 discrimination ratio were obtained from the original Wald formulas. [7] This was done because the formulas used for Test Plan III of MIL-STD-781B were themselves modifications of the original formulas. (The reject line was moved to compensate for the effects of the truncation point on the risks). While not strictly necessary, it was felt that the use of the original formulas was preferable, if for no other reasons than that the original formulas provided equations for the reject and boundary line differing only in the sign of the constant term, which simplifies both the derivation and the use of the new test plan. The equations are $T = \pm 4.394 + 1.386F$ (or $F = \pm 3.170 + .721T$).

2. For test lengths of 5, 10, 20, 40, 80 multiples of Θ_1 , new values were found for the constant which would result in approximately equal producer and consumer risks. This was done by trial and error, substituting constants into the equations and running a Monte Carlo simulation program to determine the producer and consumer risks resulting from the modified formulas. Results are presented in Table II.

Table II
Empirically Derived Constants

	Test Length (Multiples of Θ_1)				
	5	10	20	40	80
Constant	1.6	2.3	3.5	5.4	8.3
Producer Risk	.27	.21	.13	.05	.01
Consumer Risk	.25	.18	.12	.05	.01

3. From the constants shown in Table II, a 'least squares' fit was made to the formula $C = C_1 (TL) C_2$:

C function desired

C_1, C_2 constants

TL test length in multiples of Θ_1 .

The function is strictly empirical, but is tractable and represents the simplest function with a potential of satisfying the need. $C_1 = 0.594$ and $C_2 = 0.598$.

4. The Monte Carlo simulation was repeated for the same range of test lengths using constants calculated from the derived function. The results are shown in Table III.

Table III
Constants and Risks Using $C = 0.594 (TL)^{0.598}$

	Test Length (Multiples of Θ_1)				
	5	10	20	40	80
Constant	1.56	2.35	3.56	5.39	8.16
Producer Risk	.29	.21	.13	.05	.01
Consumer Risk	.27	.19	.11	.05	.01

Comparing Table III with Table II shows the least squares derived formula gives results comparable to the constants derived by trial and error. A full set of operating characteristics is provided in Fig. 3.

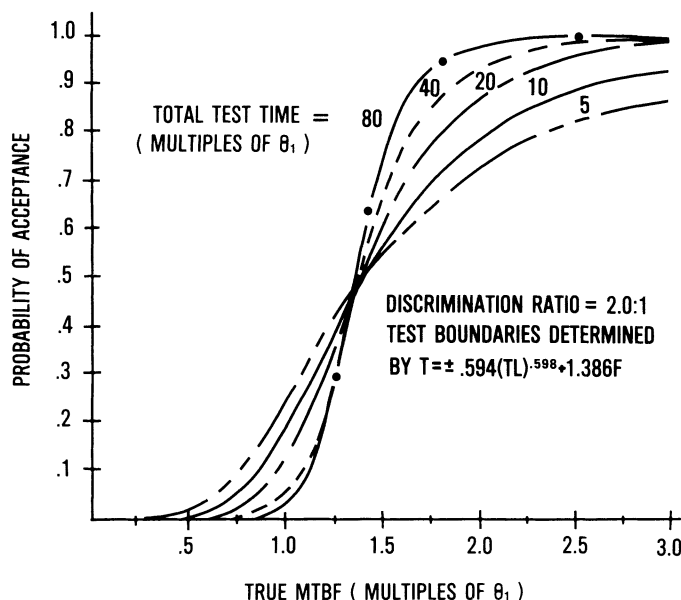


Fig. 3 Operating characteristics of modified all-equipment test plan

5. The preceding steps were then repeated to derive equations for equal risk all-equipment tests for discrimination ratios of 1.5:1 and 3.0:1. The Wald formulas for sequential tests with 10% risks were the starting points, and the final formulas were:

$T = \pm 0.673(TL)^{0.536} + 1.216F$, for a discrimination ratio of 1.5:1,

$T = \pm 0.614(TL)^{0.633} + 1.648F$, for a discrimination ratio of 3.0:1

Equivalent forms are:

$F = \pm 0.553(TL)^{0.536} + 0.455T$,

$F = \pm 0.373(TL)^{0.633} + 0.226T$, respectively

Operating characteristics for these are given in Fig. 4&5.

APPLYING THE NEW TEST

To apply the test, the user ought to select an all-equipment test based on the same discrimination ratio as the test he uses for reliability qualification. He then applies the appropriate equation from the three provided above to draw his reject and boundary lines. He must specify the test time required on each unit. From this, and the number of units produced, he can determine his risks using Figures 3–5 as appropriate. During the test, failures are plotted against time, where time is in multiples of Θ_1 . Crossing the reject line indicates a production reliability problem to be resolved. The plot is not allowed to cross the boundary line but follows the line until a failure occurs. The only difference in this procedure and the existing 2-line all-equipments test, is the use of the formulas derived above to establish the reject and boundary lines. This

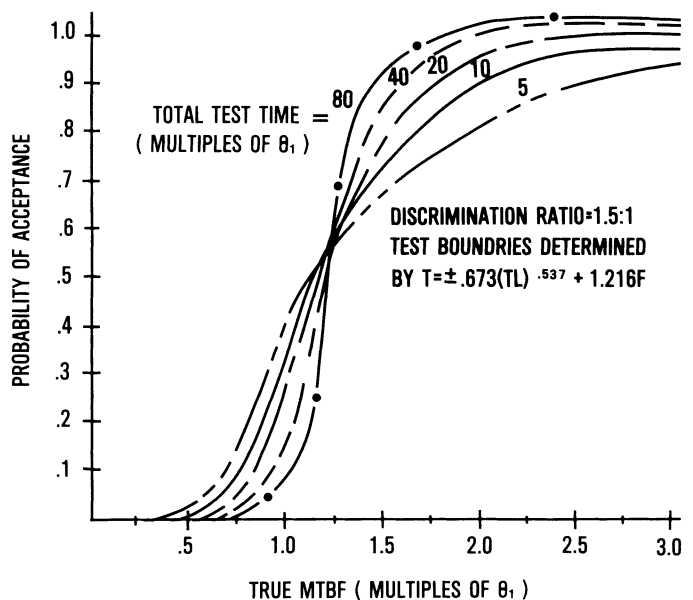


Fig. 4 Operating characteristics of modified all-equipment test plan

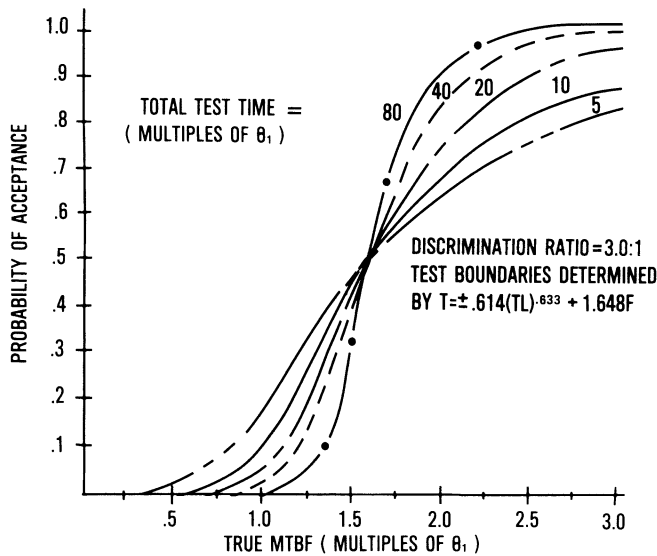


Fig. 5 Operating characteristics of modified all-equipment test plan

difference, however, keeps the producer and consumer risks approximately equal regardless of the test length.

This procedure will not, of course, solve all the difficulties with the all-equipment test. The s -risks in the shorter test periods, though now equal, can still be too high. For such cases, there is no recourse except to lengthen test time, if equal s -risks are to be preserved. On the other hand, the s -risks are very low for long tests and one may wish to reduce the test time for economic reasons, by testing every other unit, for example. The new procedure will, at least, provide that any test length selected will provide equal s -risks, preventing favoring the producer or consumer even if blindly applied by inexperienced personnel. Since not every program is blessed with trained reliability specialists, this alone could be an important benefit.

REFERENCES

- [1] R. Sesson, R. Neathammer, "An improved all-equipment reliability test plan", *Evaluation Engineering*, 1969 Sept/Oct, pp 6-8, 50.
- [2] R. Sesson, "Improved fixed length reliability qualification tests", *Evaluation Engineering*, 1972 May/Jun, pp 4, 6, 7, 40.
- [3] R. Sesson, M. Anderson, G. Lindsay, E. Thomas, "Reliability/maintainability demonstration verification", *Evaluation Engineering*, 1973 May/Jun, pp 10-13, 31.
- [4] Aeronautical Systems Division Supplement I to Air Force Regulation 80-5, 25 Sept 1970, Wright-Patterson AFB, ASC/ENES, Ohio 45433 USA.
- [5] MIL-STD-781B, "Reliability Tests: Exponential Distribution", 15 Nov 1967, Naval Electronics Systems Command, Elex 4702, Washington D.C. 20360 USA.
- [6] Draft of MIL-STD-781C, Naval Electronics Systems Command, Elex 4702, Washington D.C. 20360. USA. (Now published as a Standard and supercedes-781B.)
- [7] RADC Reliability Notebook, 31 Dec 1961, Rome Air Development Center, RBRT, Griffiss AFB NY 13441 USA, Section 5.
- [8] ARINC Research Corp, Publ. No. 123-7-196.

Anthony Coppola; RADC/RBRT; Griffiss AFB; Rome, NY 13441 USA.

Mr. Anthony Coppola (M'57, SM'77) was born in Brooklyn, New York on 1935 Jul 14. He holds a Bachelor's Degree in Physics and a Master's Degree in Engineering Administration, both from Syracuse University. He is Chief of the Reliability and Maintainability Engineering Section, Rome Air Development Center, Griffiss AFB and is responsible for the development of reliability and maintainability engineering techniques. He has published more than a dozen technical papers on the methods and management of reliability and maintainability. He has been a member of the faculty of the Air Force Institute of Technology School of Systems and Logistics and of the Staff of the George Washington University School of Engineering and Applied Science. He is a past chairman of the IEEE Mohawk Valley Section and the organizer and first chairman of the Mohawk Valley Chapter of the IEEE Reliability Group.

Manuscript received 1977 Jun 18; revised 1977 Oct 5.

□ □ □

Reply to Referee Comments & Questions

1. **Q.** The reject line of Test XXIX is most important to both the producer and consumer, so that QC difficulties can be detected and corrected early. It is not the intent of MIL-STD-781B to demonstrate reliability in Test XXIX. Rather it is to be used when inadequate QC or new production lines are involved.

A. Test Plan XXIX is not a reliability qualification test, but a production control as the referee states. However, like all statistical tests, producer and consumer risks are involved. I contend that they should be equal, presuming both are reasonable figures (i.e., the value of θ_0 is about as high as one could expect with a good, but not extra-ordinary reliability effort and θ_1 is the lowest value which is operationally and economically acceptable).

2. **Q.** Par. 4.2.9.4 indicates that the MTBF can be evaluated at any time during the contract by totalling the operating time and the failures. Corrective action will be taken in accordance with par. 5.4.8.5 when the data justify such action. The procuring agency will negotiate for the required corrective action. The producer takes no risk except the disapproval of the customer, unless penalties are part of the contract. If penalties are part of the contract, the producer will make sure that his discrimination ratio is large and that the test-time is reasonably long to minimize his risk, before signing the contract.

A. The producer cannot always "make sure his discrimination ratio is large". If he could, it would imply he can design to any value of θ_0 he desired, which is theoretically possible but can be a practical impossibility. In a competitive environment, he will not willingly commit resources above normal for reliability effort. If the request for proposal fixes the length of the test he may not wish to take exception to it. (Indeed, he may not realize what his s -risks are when writing his proposal.) Penalties are

not the only consideration. We call the risk of rejecting 'good' equipment a producer risk, but it is also against the best interest of the consumer.

3. **Q.** Test XXIX is a fixed-time as specified in the contract and is not intended for use in the Reliability Qualification (Demonstration) Phase.

A. I do not suggest using Test Plan XXIX as a qualification test. I am only trying to get it to do its function of production control without skewing the s -risks either in favor of the producer or consumer.

4. **Q.** Equal risks appear unnecessary for this test. Since the test-time is part of the contract, the producer must design sufficiently above the specified MTBF to minimize his risk of rejection if the test is contractually binding.

A. Equal s -risks should always be necessary if the Θ_0 & Θ_1 values are set realistically. If the customer has enough design margin to easily design above Θ_0 , that value has been made too low. In any event, if Θ_0 describes 'good' equipment (in the customer's opinion) it should have a low chance of rejection. I feel we must make Θ_0 & Θ_1 realistic numbers, rather than rely on the contractor having latitude to design above an arbitrary value of Θ_0 , if we are to be using an engineering discipline rather than a numbers game.

5. **Q.** The proposed test seems to be a variation of Sequential Test Plan III where total-test-time is in multiples of specified MTBF (Θ_0).

A. The new test is derived from the sequential test formulas as are Test Plan XXIX and the 2-line version in MIL-STD-781C. I am using multiples of Θ_1 because the new MIL-STD-781C will also be using multiples of Θ_1 rather than of Θ_0 as the current standard does.

6. **Q.** What purpose does the boundary line of Figure I serve?

A. The boundary line keeps the curve within a given number of failures from reject. This prevents the situation where a good equipment strays far from the reject line and then goes bad but cannot rapidly reach the reject line. The boundary line permits more rapid detection of production problems, but also can provide (in long tests) an undesirable risk of a good equipment reaching a reject point by chance. My modification to the test gives the best of both worlds by keeping a boundary line but changing its distance from the reject line sufficiently to reduce the risks of rejects caused by chance rather than by production degradation. I did not invent the boundary line. My references discuss its origin. It is used by the Air Force and is the plan in MIL-STD-781C in lieu of the present Test Plan XXIX.

7. **Q.** Figure I, if used for a reliability test, can be developed in accordance with [8] for equal risk and discrimination ratios. However, this is a Sequential Test Plan rather than a fixed-time-test.

A. Again, the sequential formulas are the starting point for the test, but the test is a fixed length test (length determined by number of units produced times time on each unit). My source of the sequential test formulas was [7], which are identical to the ARINC formulas [8].

8. **Q.** A fixed-time-test can best be evaluated at any time during the test by a point estimate from the chi-square distribution, knowing the total-test-time and observed number of failures.

A. This is irrelevant if the all-equipment test continues until it hits a reject or all equipments are produced and passed.

9. **Q.** The longer the test, the less is the probability of rejecting good equipment. Test XXIX (page 29, 781B) shows producer risk approaches 25%.

A. My discussion is not on Test Plan XXIX in MIL-STD-781B but its 2-line replacement. The producer risk in Test Plan XXIX is lower than 25% for short tests and increases as test-length increases. (See OC curves in MIL-STD-781B.)

10. **Q.** The test time is specified in the contract and is undoubtedly limited only by the consumer's funds and schedule.

A. Quite so. That is why the test should provide equal s -risks for all reasonable test periods.

11. **Q.** Are you developing a new Sequential Test Plan or a Fixed Time Test? The discrimination ratio appears to be the variable here.

A. We are developing a fixed time test with boundaries set by modifications to the sequential formulas. The variable is test length.

12. **Q.** See Figure 2. The test period (par. 4.2.9.2, -781B) can be as small as 1/4 of the specified MTBF (Θ_0). If the discrimination ratio were 2, then the test time would be $0.5 \Theta_1$. Is it true that a minimum of 10 equipments would be required here?

A. I do not recommend time and sample combinations which provide high s -risks. One would increase test-time per unit to create reasonable risks but with a small number of samples, production control could perhaps be assured by a standard test (fixed time or sequential) applied to a sample mid way in production rather than to all equipment. This is a tradeoff which the program office should make. □ □ □