ADA 278573

AFIT/GOR/ENS/94M-13

Accesion For	
NTIS CRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

AN ANALYSIS OF OPERATIONAL
SUITABILITY FOR TEST AND EVALUATION
OF HIGHLY RELIABLE SYSTEMS

THESIS

James N. Serpa, B.S.
Captain, USAF

AFIT/GOR/ENS/94M-13

DTIC
ELECTE
APR 22 1994
S G D

Approved for public release; distribution unlimited

94-12282

94 4 21 063

THESIS APPROVAL

STUDENT: Captain James N. Serpa, USAF CLASS: GOR-94M

THESIS TITLE: An Analysis of Operational Suitability for Test and Evaluation of Highly Reliable Systems

DEFENSE DATE: 4 March 1994

COMMITTEE:	NAME/DEPARTMENT	SIGNATURE
Advisor	GREGORY S. PARNELL, Col, USAF Head Department of Operational Sciences Graduate School of Engineering	GS Parnell
Reader	DENNIS C. DIETZ, Lt Col, USAF Assistant Professor of Operations Research Department of Operational Sciences Graduate School of Engineering	Dennis C Dietz

AFIT/GOR/ENS/94M-13

AN ANALYSIS OF OPERATIONAL SUITABILITY FOR TEST AND EVALUATION OF HIGHLY RELIABLE SYSTEMS

THESIS

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science

James N. Serpa, B.S.

Captain, USAF

MARCH, 1994

Approved for public release; distribution unlimited

ACKNOWLEDGMENTS

This thesis effort has shown me that no matter how much one learns, there is always more to know. Being my first large scale research effort, this experience brought out a few of my weaknesses such as poor organization and fear of failure. I hope to have learned from problems I have caused myself in this effort.

Captain Paul D'Agostino (at HQ AFOTEC/SAL) provided invaluable assistance in showing me the ropes of the test and evaluation world. Lieutenant Colonel Dennis Dietz provided very useful inputs to the Methodology and Analysis chapters -- I only wish I had involved him in the process sooner. Colonel Greg Parnell lead me through a sometimes frustrating maze of research. His insights were always eye opening for me.

My wife Kerrie agrees that AFIT was more fun before I started the thesis, and we are able to joke about it now. I am grateful to her for keeping our family, home, and finances in order when I was unable to help. Also, I am sorry for our time together that had to be sacrificed during the past few months.

James N. Serpa

Table of Contents

List of Figures

List of Tables

ABSTRACT

The purpose of this research was to develop a quantitative measure of operational suitability (OS) and determine its applicability in making the test length decision prior to Initial Operational Test and Evaluation (IOT&E). The current approach used by the Air Force Operational Test and Evaluation Center (AFOTEC) was presented and used to establish the relationships of the test measures. It was established that OS could be represented by a function of operational availability (A_0) and built-in test effectiveness (BE). BE was defined and measures proposed based on the method of data collection.

A proposal for predicting A_0, BE, and OS to determine the proper test length and sample size was analyzed for several examples of prior information. Multiplicative and additive utility functions were proposed as possible ways to calculate OS. It was shown that probability statements could be made about BE, A_0, and OS from the prior information; this analysis revealed the reliance of the results on the prior information.

AN ANALYSIS OF OPERATIONAL SUITABILITY FOR TEST AND EVALUATION OF HIGHLY RELIABLE SYSTEMS

I. Introduction

Problem Statement

The decisions made throughout the DoD acquisition process culminate with the decision to begin full-rate production of a weapon system. This decision cannot be made intelligently without knowing how well the system might perform in operational conditions. Operational Test and Evaluation (OT&E) is performed to help determine how well the system might perform in operational conditions.

Determining "how much testing is enough" has long been considered by many practitioners of the test and evaluation discipline as a classic problem (10:1). It is intuitive that more testing will help paint a more accurate picture of a system; however, there is a limit to the amount of testing that is cost effective. The cost of testing and limitations such as time and schedule constraints make it desirable to test a system only as much as is required to acquire useful results.

Test and Evaluation

The DoD acquisition process involves two basic types of test and evaluation (T&E); developmental test and evaluation (DT&E) and operational test and evaluation (OT&E). OT&E is further broken down into initial operational test and evaluation (IOT&E) and follow-on operational test and evaluation (FOT&E).

The purpose of DT&E is to demonstrate that a system design meets contractual specifications and to identify system deficiencies to the system program office (SPO). DT&E is performed by the contractor who is building the system and managed by the implementing command, usually Air Force Materiel Command (AFMC). Since it occurs in the first stages of system development, DT&E is performed using analytical models computer simulations, and limited system testing.

Figure 1.1 shows a timeline for the stages of T&E (against well known milestones in system development) and the overlapping purposes of the stages (17:41). Once DT&E is underway, IOT&E begins on system prototypes and eventually on production models of the system. Performed by the Air Force Operational Test and Evaluation Center (AFOTEC), IOT&E is expected to identify system deficiencies and tell the user what to expect from the operational system--it should be completed in time to support the full-rate production decision.

When the system is operational, the using command performs FOT&E throughout the operational lifetime of the system. FOT&E continues the focus on the user and assesses operational performance against operational criteria.

T&E is an essential part in the life cycle of systems acquired by the Air Force. Its goals at all levels of testing and evaluation include:

- Assessing and reducing risks.
- Evaluating system effectiveness and operational suitability.
- Identifying system deficiencies (7:5).

Overall, OT&E for combat systems focuses on system effectiveness and operational suitability in the combat environment. System effectiveness is the degree to which the system can accomplish its mission in field use. Operational suitability is the degree to which the system can be supported in field use.

Currently, T&E is planned using requirements established in system documentation and performed using existing DoD guidance and T&E regulations. Risk analysis is performed via the confidence interval and hypothesis testing. Sample size determination methods are based solely on the arbitrary confidence intervals of the test--they do not investigate possible benefits of additional testing.

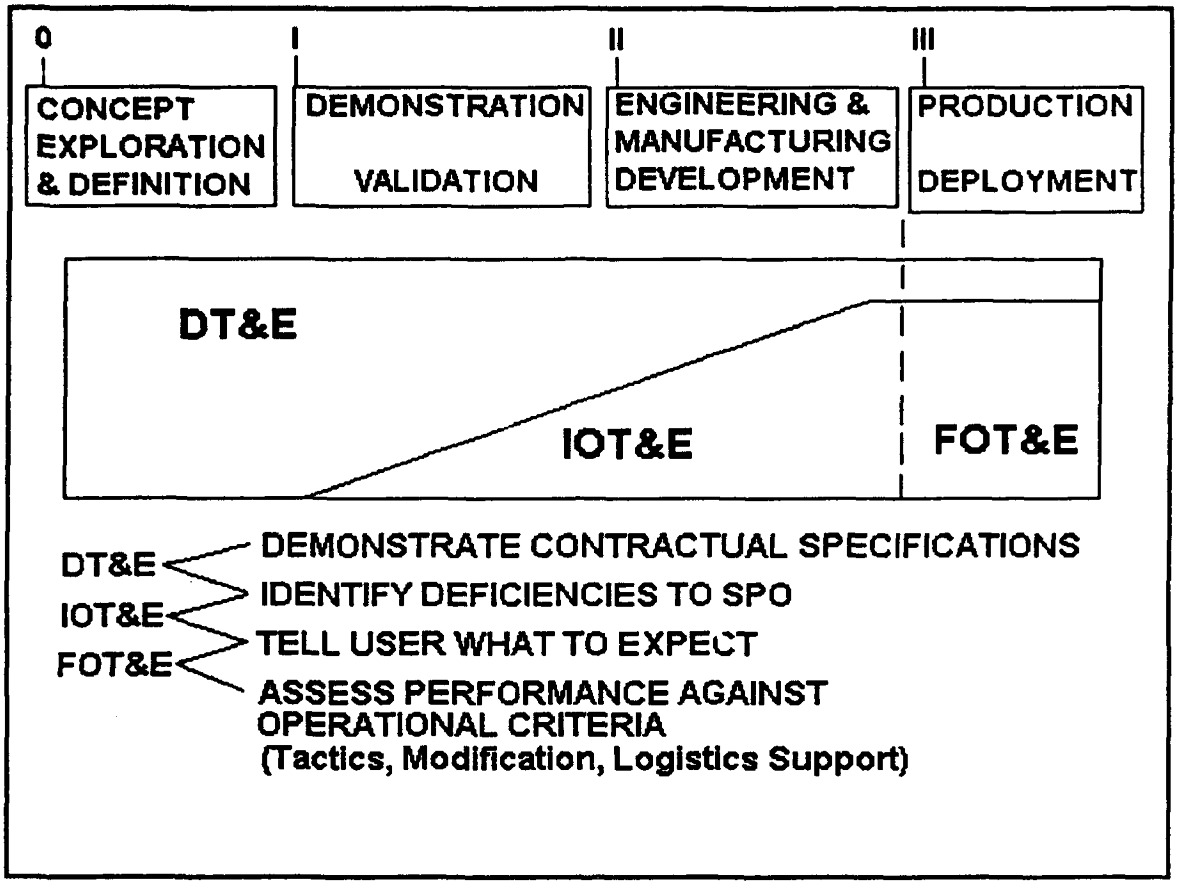


Figure 1.1 Phases of System Testing

Research Scope and Objectives

The objective of this research is to provide a method to assess the value of testing a system in support of the full-rate production decision. The research uses a suitability upgrade to electronic warfare equipment as a case study to develop and analyze the method.

The study of T&E will be limited to IOT&E and how it verifies whether a system meets operational suitability requirements. IOT&E is analyzed because it is performed to determine whether to proceed with full-rate production of the system. In addition, IOT&E has properties common to both DT&E and FOT&E, which makes it likely that the results of this research will be transferable to the other phases. Although both system effectiveness and operational suitability are evaluated in T&E, this research will focus on operational suitability testing because its methodology is better documented and its measures are common to many systems. However, the research results should also be applicable to system effectiveness testing.

Use of Decision Analysis

Decision analysis (DA) is a set of quantitative methods for analyzing decisions based on the "axioms of consistent choice" (16:356; 12:807). These axioms are simply the rules that one must adhere to in order to make consistent choices. By using DA techniques to study the testing process, we will:

- Use influence diagrams to identify the role of IOT&E in the DoD acquisition cycle and to identify the role of suitability assessments within IOT&E (1:34).
- Use probability trees to identify the relationships between built-in test (BIT) measures.
- Use stochastic analysis to apply the developed method to a scenario where the full-rate production decision is yet to be made.

Overview of Thesis

Chapter II describes the current approach to system testing employed at AFOTEC and uses two case studies to show how the approach is implemented. Also provided is a discussion of the value of the results obtained through the current approach.

Chapter III presents the development of a DA approach to sample size determination. An alternative approach is developed using one of the Chapter II case studies.

Chapter IV presents an example of how the approach developed in Chapter III can be used to determined test sample size. A stochastic analysis of the example results is presented.

Chapter V summarizes the research effort and suggests topics for further research.

II. Current IOT&E Methodology

This chapter begins with a brief overview of the IOT&E process from the perspective of AFOTEC in order to lay a foundation for presenting the current AFOTEC methodology for performing IOT&E. Two case studies are examined in order to better understand the current testing methodology and the value of the results obtained through these methods.

Overview of IOT&E

One purpose of all testing performed during the DoD acquisition process is to verify that systems are operationally effective and suitable for intended use (15:8-2). For IOT&E, this is the primary purpose because IOT&E is performed in support of the pivotal Milestone III decision (Figure 1.1). The decision whether to begin full-rate production of the system cannot be wisely made without knowledge of the system's capabilities. In order to obtain this knowledge before the system is operational, the system is tested by observing it in scenarios created to represent the system's operational environment.

DoD testing occurs in five phases: program definition, advance planning, pretest planning, execution, and reporting (8:4). During the program definition phase, the need for OT&E is determined. Once it is determined that testing is required, AFOTEC personnel become involved in planning T&E so they can focus on the most important system parameters to test.

The planning phases involve a highly iterative process of drafting and revising the documents required to build a detailed T&E plan. AFOTEC evaluates operational requirements set forth in documents such as the Mission Need Statement (MNS), the Cost and Operational Effectiveness Analysis (COEA), and the Operational Requirements

Document (ORD) (8:18). In addition to these documents, previous T&E of comparable systems is used in determining testing requirements. During the advanced planning and the pretest planning phases, AFOTEC develops the test concept by scoping the test; developing scenarios; and determining schedule, resource requirements and test limitations (8:18).

It is during these early phases that Critical Operating Issues (COIs) are determined and from them measures of effectiveness (MOEs) and measures of performance (MOPs) are derived (8:19). These measures and the requirements established in the Operational Requirements Document (ORD) are the basis for AFOTEC's test criteria (8:10). All source documents developed in these preparatory phases are coordinated in the Test and Evaluation Master Plan. Together, these documents provide the framework, basic test philosophy, and guidance required to build a detailed OT&E plan (8:10).

AFOTEC's involvement in the pretest planning, test execution, and reporting phases is outlined in AFOTECI 99-101, chapters 3, 4, and 5, respectively. Throughout test execution, the AFOTEC test team ensures the correct data is being properly collected (8:43). Prior to the end of testing, all data must be aggregated and analyzed for the final report.

Suitability Testing. As previously mentioned, IOT&E is comprised of operational suitability testing and system effectiveness testing. The objective of operational suitability IOT&E is to ensure that new systems can be operated and maintained in field conditions (9:Chapter 2,1). The objective of system effectiveness IOT&E is to ensure that new systems can effectively perform the missions for which they were designed (17:42). Whereas both of these elements of IOT&E are important and have commonalities, they also have unique characteristics that require they be treated separately. This research emphasizes the elements of suitability testing while mentioning applicability to effectiveness testing when appropriate.

In planning IOT&E, considerations are taken so the data required to make suitability and effectiveness determinations can be obtained. The test results, often a combination of quantitative and qualitative data, must be combined in such a way as to show whether the system is effective and suitable. The methods used to prepare suitability data are presented in these sections.

A recurrent theme in all guidance documentation concerning operational suitability is the need to evaluate the system's ability to meet operational readiness requirements (9:Chapter 9,1). Availability, reliability and maintainability of the system are dominant factors in determining whether the system is operationally suitable. AFOTECP 400-1, Part III, Chapters 9, 10, and 11 detail these factors to include how they are measured and how they are interrelated.

Case Studies

Two systems, each involving a unique suitability testing situation, are used as case studies for this research. These systems were chosen because they are representative of systems that can be treated as a "black box." This is important because it provides for the use of widely accepted assumptions about their performance (9:Chapter 10,5). Also, the systems are examples where the information that can be obtained with a reasonable amount of testing varies greatly.

AN/ALR-69 Radar Warning Receiver (RWR). This system is a reliability and maintainability (R&M) modification package designed to avoid future supportability problems caused by vanishing sources of supply and obsolescence of existing system components (3:1). Because the sole purpose of this modification is to improve system R&M, the effectiveness testing requirements are not typical of most systems. The effectiveness requirements are only that the modified system be at least as effective as the existing system. Basically, the R&M improvement may not detract from the current operational effectiveness of the system.

Suitability. The methodology of the test is to perform analysis based on point estimates from the observed test data (3:2). The test must answer the Critical Operating Issue (COI): Will the R&M modification to the AN/ALR-69 RWR maintain operational reliability and improve maintainability and availability to support mission accomplishment? This COI is supported by nine MOPs, which are used to measure the operational suitability for the system (3:3). Five of the MOPs relate directly to testing the system's Built-in Test (BIT) capability and Integrated Diagnostics Effectiveness (ID). In addition to the data that will be collected during the test, a maintenance demonstration

Table 2.1 AN/ALR-69 RWR Measures, Criterion, and Test Results

Measures of Performance (MOPs)	Criterion	Test Results
Operational Availability (A_0)	95 %	100 %
Mean Time Between Critical Failure (MTBCF)	42.3 hours	≥ 232 hours
Mean Repair Time (MRT)	1.75 hours	not observed
Mean Down Time (MDT)	2.25 hours	not observed
Integrated Diagnostics Effectiveness (ID)	100 %	100 %
BIT Fault Detection Rate (FDR)	90 %	100 %
BIT Fault Isolation Rate (FIR)	90 %	100 %
BIT False Alarm Rate (FAR)	5 %	0 %
BIT Fault Detection Time (FDT)	Operator BIT: 45 seconds Mx BIT: 180 seconds	~ 0 seconds ~ 0 seconds

(M-Demo) will be performed to evaluate BIT and ID capability. The remaining four MOPs include: Operational Availability (A_0), Mean Time Between Critical Failure (MTBCF), Mean Repair Time (MRT), and Mean Down Time (MDT). Each of the MOPs has a criterion that will be used to measure the performance of the system during the test. These criteria and the actual results of the test, which was completed in August 1993, are contained in Table 2.1. Of the two case studies, the AN/ALR-69 RWR is unique in that it is the only one in which complete test results are available.

For this system, the number of critical failures experienced during the test was used to determine potential confidence in the MTBCF measurement. A standard test confidence requirement for all T&E is 0.80, or *P[type-I error]* $\leq$ *0.20.* The confidence level for *MTBCF = 42.3 hours* was used to set the number of test hours required because MTBCF was determined by AFOTEC to be the guiding MOP for the AN/ALR-69 RWR suitability test (3:2). Potential confidences for this test, which depend on the number of critical failures observed and the number of test hours, are shown in a Confidence Table (see Table 2.2).

Table 2.2 Confidence Table for the AN/ALR-69 RWR

\Test hours Critical failures\	180	200	220
0	0.99	0.99	0.99
1	0.93	0.95	0.97
2	0.80	0.85	0.89
3	0.61	0.69	0.76

Confidence tables are constructed using a standard one-tailed confidence interval for a time-terminated test (Equation (2-1)), where T is the length of the test in hours, θ is the desired MTBCF in hours, R is the number of critical failures, and α is the P[type-I error] of the χ^2 statistic with 2*R+2 degrees of freedom (11:254).

$$\frac{2*T}{\chi^2_{\alpha,2*R+2}} \leq \theta = MTBCF \qquad (2\text{-}1)$$

The table values are (1 - α) for the value of α that solves the equality form of Equation 2-1. For example, if two critical failures are experienced during a 200 hour test, then there will be 0.85 confidence that MTBCF > 42.3. From this confidence table, as many as two critical errors can be experienced during a 180 to 220 hour test and the test results would still have the required 0.80 confidence level.

Military Microwave Landing System Avionics (MMLSA). The MMLSA provides a single all-weather precision approach and landing aid operable with military, civil, and international bases (4:1). IOT&E is scheduled for September 1995 on a production representative MMLSA maintained by USAF personnel.

Suitability. The methodology of the test is to perform analysis based on mathematical models and observed test data (4:2). The test must answer two critical operating issues; COI-1: Does the MMLSA have adequate design and reliability to support worldwide deployment? and COI-2: Does the MMLSA maintenance fully support user mission requirements (4:2)?

COI-1 is supported by three measures: MTBCF, mean time between corrective maintenance action (MTBCMA), and on-equipment mean repair time (MRT) (4:3). These measures present important risk-reduction methods for testing: recently developed design of experiments (DOE) tables and the use of Bayesian statistics to increase the confidence level of the test results that will be obtained from a small sample size (4:2). For the Bayesian analysis, data from developmental test and evaluation (DT&E) is used to form the prior estimates. The prior will then be updated with the current test data to get the final results in the form of the posterior. This method is chosen because the expected values for the MTBCF measure is much greater than the expected number of test hours.

COI-2 is supported by two measures: Integrated Diagnostics Effectiveness (ID) and Fault Detection Rate (FDR). In addition to the data that will be collected during the

Table 2.3 MMLSA Measures and Criterion

Measures of Performance/Evaluation (MOPs/MOEs)	Criterion
Mean Time Between Critical Failure (MTBCF)	2300 hours
Mean Time Between Corrective Maintenance Action (MTBCMA)	2000 hours
Mean Repair Time (MRT)	0.26 hours
Integrated Diagnostics Effectiveness (ID)	100 %
BIT Fault Detection Rate (FDR): for critical failure to 1 LRU	99 %
for failure to 1 LRU	85 %

test, a maintenance demonstration (M-Demo) will be performed to evaluate BIT and ID capability. The criteria for each measure are summarized in Table 2.3. Since IOT&E has not been performed and DT&E is not complete, there are no results to show at this time.

Table 2.4 Confidence Table for MMLSA

\Test hours Critical failures\	172	192	212
0	0.06	0.07	0.08
1	0.00	0.00	0.00

A confidence table for MMLSA (Table 2.4) is also built using the MTBCF measurement. Notice that the confidences are much lower for this test than in Table 2.2. This shows the need for the use of Bayesian methods to increase the number of test hours from the current allocation of 192 hours. Even with the additional test hours (number of hours is unknown at this time) provided from the DT&E data, it is unlikely the required 0.80 confidence level will be achievable due to the large criterion for MTBCF (2300 hours). Using Equation (2-1), approximately 3500 test hours would be required with zero failures in order to have a 0.80 confidence that the MTBCF criterion is met (6:1).

Value of Operational Suitability IOT&E Results

Given that the purpose of IOT&E is to support the full-rate production decision for a system, an appropriate measurement of value of IOT&E results would be the amount of additional information the test results provides. This section discusses the statistical value of the actual test results for the AN/ALR-69 RWR and the hypothetical test results for the MMLSA.

AN/ALR-69 RWR. The results of suitability IOT&E for this system, shown in Table 2.1, are representative of test results for a highly reliable system. In 232 hours of testing, there were no system critical failures. An extension of Table 2.2 would show that confidence for the requirement *MTBCF = 42.3 hours* is greater than 0.99; this result implies the actual MTBCF is probably much greater than the desired 42.3 hours. Through the confidence table, one is able to see how different test results (i.e., one critical failure is observed) would change the test confidence.

When no repairs are required, MRT and MDT cannot be computed; both can be assumed equal to zero or considered "not observed."

$$A_o = \frac{MTBCF}{MTBCF + MDT} \qquad (2\text{-}2)$$

Similarly, Operational Availability (A_O), which is computed using Equation (2-2), cannot be calculated since it is not possible to calculate $MTBCF = Test\ Time\ /\ \#\ of\ critical\ failures$ when there are no critical failures. Currently, AFOTEC assumes A_O=1.0 when this situation arises since the availability is 100 % for the duration of the test (no down-time). The measures for BIT capability have a limitation in that there were no problems experienced during the test or during the maintenance demonstration.

These observations point out some difficulties that arise when testing a highly reliable system. Since the requirement for the primary measure, MTBCF, was only 42.3 hours, the test time of 232 hours and the zero critical failures were useful in providing a high confidence that the system meets the requirement.

MMLSA. Since IOT&E has not yet been performed on this system, it is only possible to speculate as to the value of potential results. As is shown in Table 2.4, the expected test length will not produce a 0.80 confidence level. Even with the use of DT&E results to increase the test sample size, the confidence level will be far from 0.80. Further analysis of this system is deferred to Chapter V, where further research is suggested.

Summary

Procedures have been developed to assess operational suitability requirements during IOT&E. However, constraints such as cost and schedule often limit the amount of testing that can be performed. The impact of limited test length or sample size on the confidence level of the test results varies by system. The expected test confidence level for the AN/ALR-69, a highly reliable system with a relatively low test requirements, is sufficient despite test constraints (Table 2.2). However, the expected test confidence level for the MMLSA, a highly reliable system with a relatively high test requirements, is not sufficient (Table 2.4).

There is no single measure of operational suitability that can be used to assess the value of additional testing (increased test length or sample size). Such a measure is introduced in the next chapter.

III. Methodology

This chapter presents a method to quantify the operational suitability of a system being tested. The goal is to use the relationship between test length (or sample size) and the parameters being tested to derive a quantitative measure of operational suitability. The measure will facilitate analysis of the effect test length (or sample size) has on a system's operational suitability.

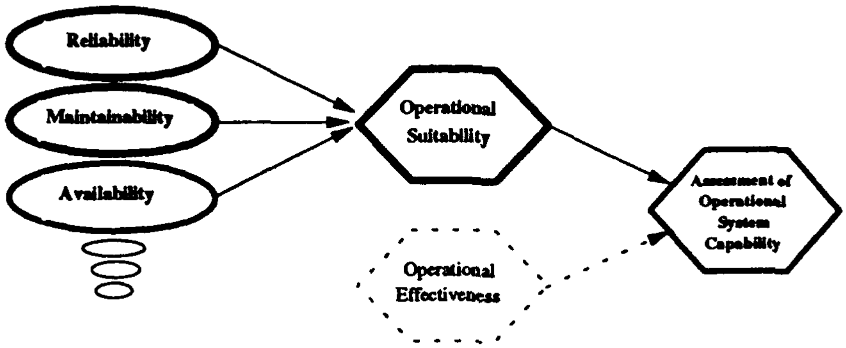


Figure 3.1 Suitability Components of IOT&E

Development

The suitability and effectiveness testing components of IOT&E and the main sub-components of suitability testing are shown in Figure 3.1. Reliability, maintainability, and availability (RMA) are the primary contributors to an assessment of suitability (2:1-1). The AN/ALR-69 RWR case study presented in Chapter 2 is used to analyze the suitability component and its sub components.

An influence diagram shows how suitability is broken down into MOPs and test parameters for this system (Figure 3.2). This influence diagram has been annotated to show how the test data could be used to calculate a measure of suitability. Above each

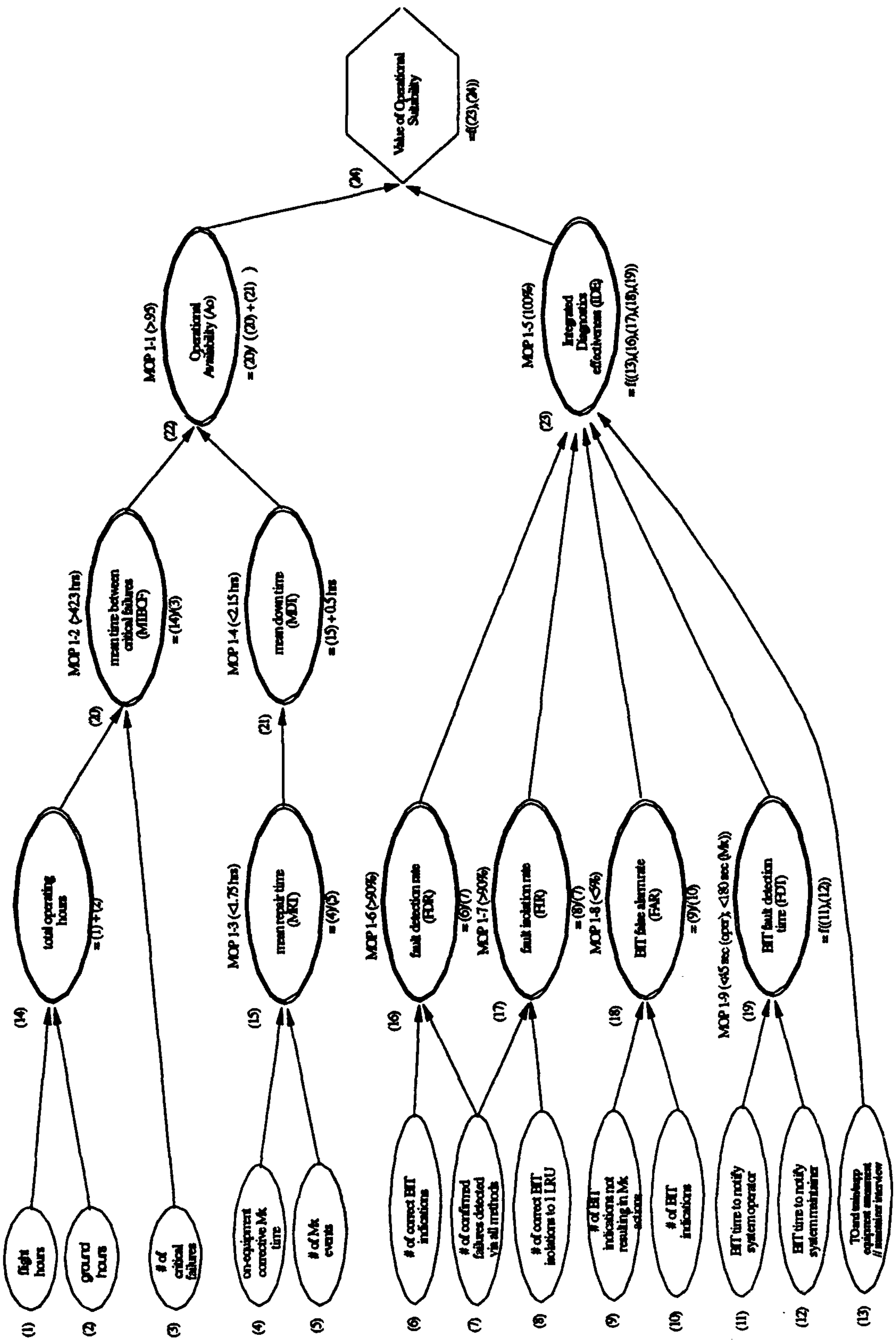


Figure 3.2 Influence Diagram of AN/ALR-69 Suitability

applicable node is the MOP number and test requirement. Nodes 1 to 13 are observed during testing and the remaining nodes are calculated. Below nodes 14 to 22 are the equations used to calculate the measure for each node, where applicable. Below node 23, which is not currently calculated quantitatively, a functional relationship is shown. The same is true for node 24, which represents a quantitative measure of operational suitability. It is evident from Figure 3.2 that operational availability (A_0) and ID Effectiveness (IDE) are direct contributors to a measure of suitability; therefore, a quantifiable measure of suitability will be a function of A_0 and IDE (Equation (3-1)). It is unnecessary to include the other MOPs directly in Equation (3-1) since they are already represented in the A_0 and IDE calculations.

$$Suitability = f(A_o, IDE) \qquad (3\text{-}1)$$

Assumptions. RMA MOPs are quantitative in nature, which aids in using them to measure suitability. The remaining suitability components, such as interoperability and compatibility, and the MOPs used to measure them are qualitative in nature. Their use in determining a quantitative measurement of suitability would produce a result with arbitrary scale and little additional significance. In addition, these components and their measures typically play a minor role in determining system suitability when used with the RMA measures (5).

The IDE MOP is a good example of this. This MOP is determined by BIT MOPs and a qualitative assessment of technical orders (TOs) and training/support equipment. Because the BIT MOPs heavily outweigh the other assessments, they are used as the sole determiner of IDE (5). For the same reason, the BIT 'time to notify the operator/maintainer' is deleted from these calculations. Figure 3.2 is simplified by these assumptions to produce Figure 3.3, in which IDE is now determined solely by BIT

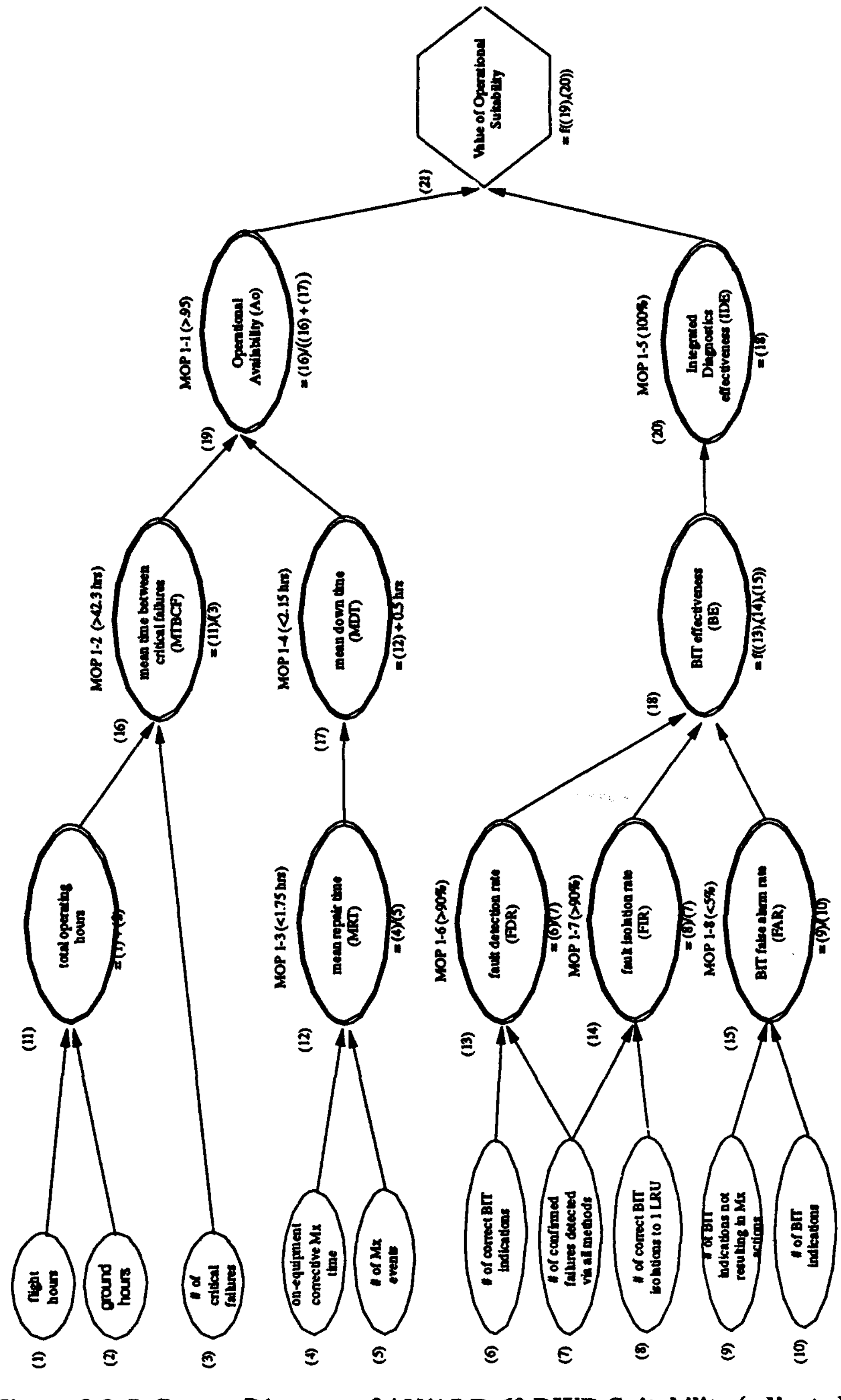


Figure 3.3 Influence Diagram of AN/ALR-69 RWR Suitability (adjusted)

effectiveness. BIT effectiveness and several approaches to calculate it are presented in the next section. Potential test requirements for this measure are discussed in Chapter IV.

BIT Effectiveness. This research uses the term BIT effectiveness (BE) to measure the capability of the BIT system. BE is defined as the probability of the BIT system making a correct decision. A correct decision is detecting and isolating a fault when it occurs and not detecting a fault when it does not occur. This general definition can be tailored as required for use in different testing situations.

Equation (3-2) defines the BIT MOPs fault detection rate (FDR), fault isolation rate (FIR), and false alarm rate (FAR) using conditional probabilities. The probability tree in Figure 3.5 shows their relationship to each other graphically. The three events described in Figure 3.5 are 1) failure occurs, 2) failure is detected by the BIT, and 3) failure is isolated by the BIT.

$$\begin{aligned} d &= FDR = P[D = 1 | F = 1] \\ i &= FIR = P[I = 1 | D = 1, F = 1] \\ FAR &= P[D = 1 | F = 0] \end{aligned} \tag{3-2}$$

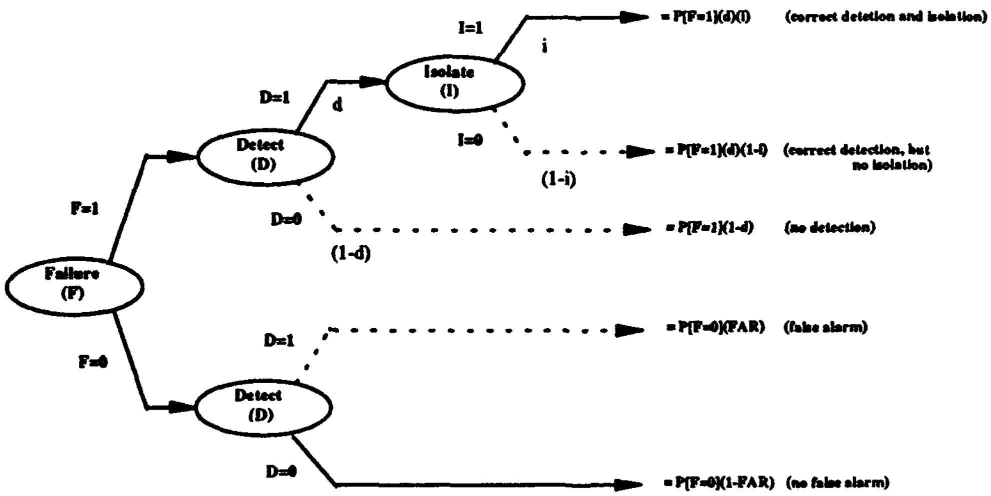


Figure 3.4 Probability Tree for AN/ALR-69 RWR BIT Effectiveness

Equation (3-3) calculates BE as a function of the BIT MOPs. However, this equation assumes the test is composed of a single Bernoulli event with a probability of false alarm, failure, detection, and isolation occurring in the event. This is not the case for the AN/ALR-69 RWR flight/ground test or maintenance demonstration (M-Demo), but Equation (3-3) provides a basis for calculating BE as a function of the BIT MOPs in each case.

$$BE = P[F=0](1-FAR) + P[F=1](di) \tag{3-3}$$

BE Observed During a Time-terminated Test. The AN/ALR-69 RWR flight/ground test is a time-terminated test -- a test in which test data is collected until a predetermined amount of time elapses. When it is suspected a system will fail a sufficient number of times during a test to adequately evaluate BIT capability, then BIT performance data is collected during the test and a M-Demo is not required.

The Poisson distribution is commonly used to model the outcomes of continuous-time tests (2:5-8). Since it is assumed for the AN/ALR-69 RWR flight/ground test that failures occur one at a time and that the number of failures and false alarms is related directly to the amount of test time, the test is modeled as a Poisson process with constant failure and false alarm rates. It follows that for F>0, $d=FDR$ and $i=FIR$ are constant and independent of the number of failures F. The detection (or isolation) of each failure is a Bernoulli event, as described earlier, so detections (or isolations) are modeled using the binomial distribution.

The next section presents two approaches to analyze BE when it is observed during a time-terminated test. The first approach is used with and without the occurance of false alarms during the test (FAR=0 or 0≤FAR≤1) and the second approach is used only when false alarms cannot occur during the test (FAR=0).

Approach #1. Due to the complexity of calculating BE during a time-terminated test, this approach uses a more restrictive definition of BE than previously presented. For this approach, BE is defined as

*BE(t) = P[no false alarms during time (t)]*P[n failures during time (t)]*
**P[n failures are detected and isolated|n failures during time (t)]*

Equation (3-4) is used to predict BE where λ = failure rate, t = test length, n = number of observed failures, and μ = FAR. When it is assumed that false alarms do not occur during the test, μ=0 in Equation (3-4), resulting in Equation (3-5).

$$\begin{aligned} BE(t) &= e^{-\mu t}\sum_{n=0}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{n!}(di)^n \\ &= e^{-\mu t}e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(di\lambda t)^n}{n!} \\ &= e^{-\mu t-\lambda t+di\lambda t}\sum_{n=0}^{\infty}\frac{e^{-di\lambda t}(di\lambda t)^n}{n!} \\ &= e^{-[\mu+(1-di)\lambda]t} \qquad (3\text{-}4) \end{aligned}$$

$$BE(t) = e^{-(1-di)\lambda t} \qquad (3\text{-}5)$$

As test length increases from 0 to ∞, BE decreases from 1 to 0 exponentially regardless of the inclusion of false alarms in the calculation because the larger the test length, the lower the probability of perfect performance. The term $\mu+(1-di)\lambda$ is the rate at which false alarms and undetected (or isolated) failures occur.

Approach #2. Whereas approach #1 is based on the probability the BIT detects and isolates all failures for the length of the test, this approach is based on the proportion of failures the BIT detects and isolates. The difference is that this approach gives credit for detecting and isolating one or more of the failures while the approach #1 only gives credit for detecting and isolating all failures. This approach assumes FAR≈0, which is a realistic assumption for many systems.

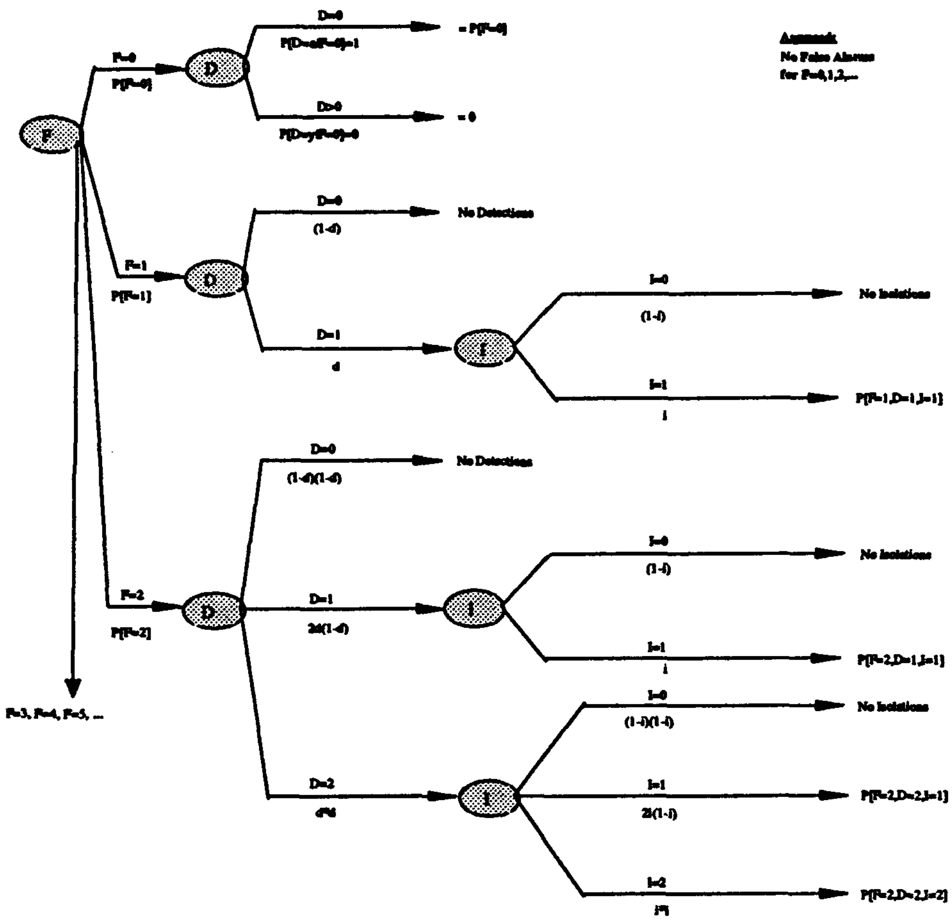


Figure 3.5 Probability Tree for Approach #2

Figure 3.5 is a probability tree that models BE for this approach. The tree is enumerated for the cases where 0, 1, or 2 failures are observed during the test. When the F=0 and F=1 cases are viewed together, the tree is identical to Figure 3.5. When two or more failures are possible, the BE prediction is complicated by the possibility of partial BE (e.g. *F=3,*

D=2, and I=1). 'F=3, D=2, and I=1' is the case where three failures occur, two of the failures are detected, and one of the failures is isolated. Since detection and isolation are binomial, there are $\binom{3}{2}=3$ ways to detect two of three failures and $\binom{2}{1}=2$ ways to isolate one of two detected failures. This results in 6 ways for the BIT system to detect and isolate one failure when three failures occur. Since the BIT is totally correct 1/3 of the time, the F=3, D=2, I=1 term is multiplied by 1/3 in the BE calculation. This is generalized to weight each term with the coefficient I / F. Equation (3-6) shows BE for any number of failures (F), detections (D), and isolations (I) as proposed in Figure 3.5. By manipulating the variables D and I to form binomial distributions that are summed from 0 to ∞ (and therefore equal to 1), Equation (3-6) simplifies to Equation (3-7), which is used to predict BE. As test length goes from 0 to ∞, BE will go from 1 to (di) exponentially. λ will determine the rate at which BE approaches (di).

$$BE(t)=e^{-\lambda t}+\sum_{F=1}^{\infty}\frac{(\lambda t)^F e^{-\lambda t}}{F!}\sum_{D=1}^{F}\sum_{I=1}^{D}\binom{F}{D}d^D(1-d)^{(F-D)}\binom{D}{I}i^I(1-i)^{(D-I)}\frac{I}{F} \tag{3-6}$$

$$=e^{-\lambda t}+\sum_{F=1}^{\infty}\frac{(\lambda t)^F e^{-\lambda t}}{F!}\sum_{D=1}^{F}\binom{F}{D}d^D(1-d)^{(F-D)}\frac{1}{F}\sum_{I=1}^{D}\binom{D}{I}i^I(1-i)^{(D-I)}I$$

$$=e^{-\lambda t}+\sum_{F=1}^{\infty}\frac{(\lambda t)^F e^{-\lambda t}}{F!}\sum_{D=1}^{F}\frac{F!}{D!(F-D)!}d^D(1-d)^{(F-D)}\frac{1}{F}\sum_{I=1}^{D}\frac{D!}{I!(D-I)!}i^I(1-i)^{(D-I)}I$$

$$=e^{-\lambda t}+\sum_{F=1}^{\infty}\frac{(\lambda t)^F e^{-\lambda t}}{F!}\sum_{D=1}^{F}\frac{(F-1)!}{D!(F-D)!}d^D(1-d)^{(F-D)}\sum_{I=1}^{D}\frac{D!}{(I-1)!(D-I)!}i^I(1-i)^{(D-I)}$$

Let q = I-1 and r = D-1.

$$=e^{-\lambda t}+\sum_{F=1}^{\infty}\frac{(\lambda t)^F e^{-\lambda t}}{F!}\sum_{D=1}^{F}\frac{(F-1)!}{k!(F-D)!}d^D(1-d)^{(F-D)}iD\sum_{q=o}^{r}\frac{r!}{(q)!(r-q)!}i^q(1-i)^{(r-q)}$$

Let s = F-1 and t = D-1.

$$= e^{-\lambda t} + \sum_{F=1}^{\infty} \frac{(\lambda t)^F e^{-\lambda t}}{F!} di \sum_{t=0}^{s} \frac{(s)!}{t!(s-t)!} d^t (1-d)^{(s-t)} (1)$$

$$= e^{-\lambda t} + \sum_{j=1}^{\infty} \frac{(\lambda t)^j e^{-\lambda t}}{j!} di(1)(1)$$

$$BE(t) = e^{-\lambda t} + (1 - e^{-\lambda t})(di) \tag{3-7}$$

A characteristic of Equation (3-7) is $BE(t) \approx di$ for large test lengths. This characteristic leads to the development of a BE measure for when a large test length is desired, but a shorter test is performed due to contraints such as cost and schedule.

BE Observed During M-Demo. Whereas operational availability can be always be observed during the flight/ground test, the capability of the BIT system cannot be observed when a system does not fail during the test. There can be a very high probability of zero failures during a test when a system is highly reliable. For the AN/ALR-69 RWR, the predicted bench reliability of 10,000 hours means few or no failures are expected to occur during a test of a few hundred hours. In response to this problem, the current AFOTEC policy is to perform a M-Demo so BIT capability can be observed. The M-Demo is made up of a number of system faults presented one at a time to a system maintainer. This gives a tester the ability to observe the BIT detect and isolate the induced faults.

BE can be predicted for a M-Demo by using Equation (3-8), which is Equation (3-7) with large test length ($t \rightarrow \infty$). These assumptions are logical since a M-Demo must be made up of at least 1 failure and for a highly reliable system, a large test length would be required to experience a large number of failures. BE can be predicted and calculated to provide a quantifiable input into an operational suitability measure.

$$BE = di \tag{3-8}$$

Operational Availability. A_O is observable for any test length since it is assumed perfect when there are no failures and calculated using Equation (3-9) otherwise. MTBCF, MRT, and MDT affect the measure of suitability through A_O and therefore need not appear directly in the operational suitability equation (3:3). (note: *MDT=MRT+0.5hrs*)

$$A_o = f(MTBCF, MDT) = \frac{MTBCF}{MTBCF+MDT} \tag{3-9}$$

Operational Suitability. This research presents a quantitative measure of operational suitability as a function of A_O and BE. The use of multiplicative versus additive utility functions, which produce quadratic and linear suitability indifference curves, respectively, is explored. Each type of function has underlying assumptions that are examined. Regardless of the type of utility function used, a quantitative measure of operational suitability should:

- accurately represent the importance of the function variables.
- convey the level of operational suitability and its relative meaning.

Multiplicative Utility Function. Equation (3-10) shows suitability calculated using the Cobb-Douglas (CD) multiplicative function (13:91). The exponents show the importance of each variable in determining the system suitability. Since both A_O and BE range from 0 to 1, use of this type of function gives suitability the same range of 0 to 1. The CD function will result in identically-shaped (homothetic) indifference curves -- the functional relationship is consistent for all levels of suitability (13:92).

$$Suitability = BE^{\alpha} \cdot A_o^{(1-\alpha)} \tag{3-10}$$

Additive Utility Function. Equation (3-11) shows suitability calculated using a linear additive utility function (13:93). As with the exponents in the CD function, the coefficients in the additive function show the importance of A_o and BE in determining system suitability. Since both A_o and BE range from 0 to 1, use of this type of function gives suitability the same range of 0 to 1 only if the coefficients add to 1. The additive function, which implies A_o and BE are "perfect substitutes," will result in linear indifference curves (13:93). Regardless of the type of utility function used, the weighting factors are system dependent. Without any knowledge of the proper weighting, a naive weighting in which $\alpha = 0.5$ is appropriate because it weighs each variable equally in the calculation.

$$Suitability = \alpha \cdot BE + (1-\alpha) \cdot A_o \tag{3-11}$$

Summary

In this chapter, operational suitability was defined quantitatively as a function of BE and A_o -- the multiplicative and additive utility functions were proposed for this measure. Two approaches were presented to quantify BE during a time-terminated test, the second approach resulting in the M-Demo measure of BE.

In Chapter IV, the methodology presented in this chapter is analyzed for several examples of system and BIT performance. In the analysis, BE is calculated using approach #2 (as adapted for a M-Demo) and operational suitability is calculated using the multiplicative function.

IV. Analysis of Results

When prior test results and knowledge of system performance are available and indicate that a system is highly reliable, this information can be used to minimize the amount of testing performed on the system to show that it meets operational requirements. This chapter presents an analysis of the methods (developed in Chapter III) in which BE, A_O, and operational suitability are calculated to assist in determining the appropriate test length for a highly reliable system.

Although IOT&E is complete for the AN/ALR-69 RWR, this highly reliable system and its test requirements are used as though IOT&E has not been performed. This is a situation where a M-Demo would be used to measure BE. Recall that for a M-Demo, BE is defined as the probability that the BIT system detects and isolates a failure. M-Demo will be used to measure BE. Approach #1 to calculate BE during the flight/ground test is not analyzed but revisited in Chapter V a possible area for further research.

For this analysis, the M-Demo sample size and the flight/ground test length are the unknown variables of interest. By identifying the impact of these variables on BE, A_O, and OS measures, a knowledgable test length/sample size decision can be made. In addition to assisting the pre-test test length/sample size decision, it is shown how these equations can be used to assess BE, A_O, and OS from the test data.

Review of Current Test Size Determination Method

The current method (confidence intervals) used to determine the appropriate test length for the AN/ALR-69 RWR flight/ground test was presented in Chapter II. Table 2.2 was built using this method, which assumes exponential time between failures. It showed that for a test length of approximately 200 hours, as many as 2 failures could be observed and the test confidence level for the MTBCF measure would still be over 0.80.

The current sample size determination approach for the M-Demo uses a sample size of approximately 25 when system being tested is mature. Failures presented in a M-Demo are not distributed as they actually occur, but uniformly for two main reasons: 1) the distribution of BIT detectable failures is not usually known at this point in the acquisition process, and 2) given the first reason, to ensure all likely failures are presented. The M-Demo is structured with the knowledge that the resulting data is not a representative random sample of the entire failure population and cannot be used to make statistically valid predictions of future performance (2:Chapter 3, 12). With this caveat, the data from M-Demos is used as a 'best estimate' for this research.

BE Observed During a M-Demo

As previously mentioned, a M-Demo is performed when it is believed a system may not fail more than a few times during its test. If any BIT detectable failures do occur during the test, they are simply added to the M-Demo results. For the M-Demo analysis, it is assumed that no failures occur during the flight/ground test.

Information on previous BIT performance for similar systems could be used to predict the BE for the M-Demo. For example, if BIT systems in the past have detected (or isolated) 99% of all faults, then this information can be used to form a prior distribution for d (or i). The prior system performance can then be used to predict the range and mean of BE for M-Demos of various sample sizes. By comparing predicted BE performance with its known test requirement, the appropriate sample size can be chosen.

Predicting the Range and Mean of BE. This section presents a method to estimate the BE distribution and mean. For a sample size (F), Equation (4-1) is used to determine the values of the BE probability mass function (pmf).

$$P[BE = x] = \sum_{D=I}^{F} \binom{F}{D} d^D (1-d)^{(F-D)} \binom{D}{I} i^I (1-i)^{(D-I)} \tag{4-1}$$

The cumulative distribution function (cdf), which can be created from the pdf, is used to make a probability statement about BE for various sample sizes. While either of these functions could be used to study the effect of sample size on BE, the cdf is used because it directly translates to a probabilty statement of BE. For instance, the cdf shows the probability BE meets the test requirement given a specific level of BE performance and a specific M-Demo sample size (Figure 4.1). The figures used to show cdfs are actually step functions -- the software used to generate them cannot create readable step functions for multiple functions. (note: BE test requirement is $x=0.81$)

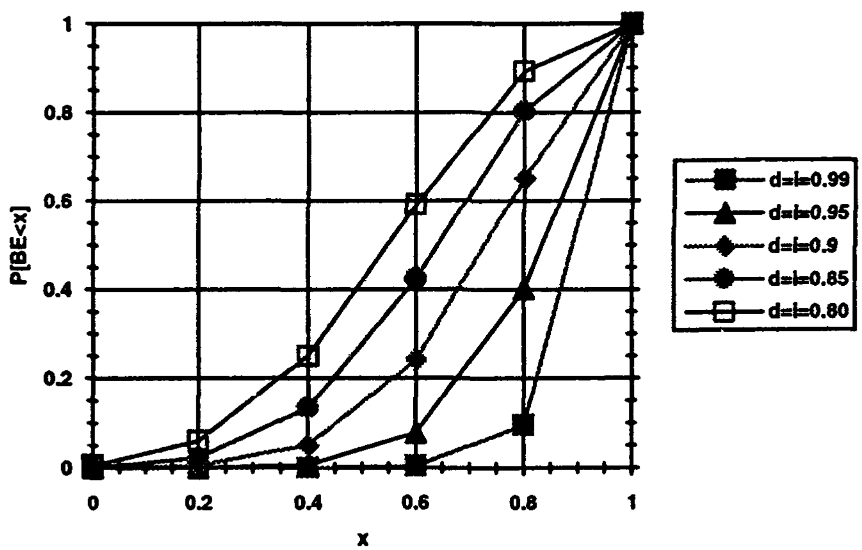


Figure 4.1 BE CDF (Sample Size=5)

If d and i are predicted to be 0.95 from prior information and the M-Demo sample size is 5, then the $P[BE \geq 0.80] \approx 0.92$ -- this is probability of meeting the test requirement if 0.80 ≈0.81 is acceptable). For any d and i, $BE\ mean = di$ regardless of the sample size. Whereas increasing the sample size from 15 to 25 does not affect the predicted BE mean, Figures 4.2 and 4.3 show it does decrease the variability of BE values.

When sample size = 15 and $d=i=0.95$, $P[BE \geq 0.80] \approx 0.95$ (Figure 4.2). When *sample size* = 25 and $d=i=0.95$, $P[BE \geq 0.80] \approx 0.99$ (Figure 4.3). The decreasing variability of BE indicates a more accurate test at higher sample sizes. Note that lower predicted *d* and *i* values make it necessary to use a larger sample size to minimize the probability of a good system failing the test.

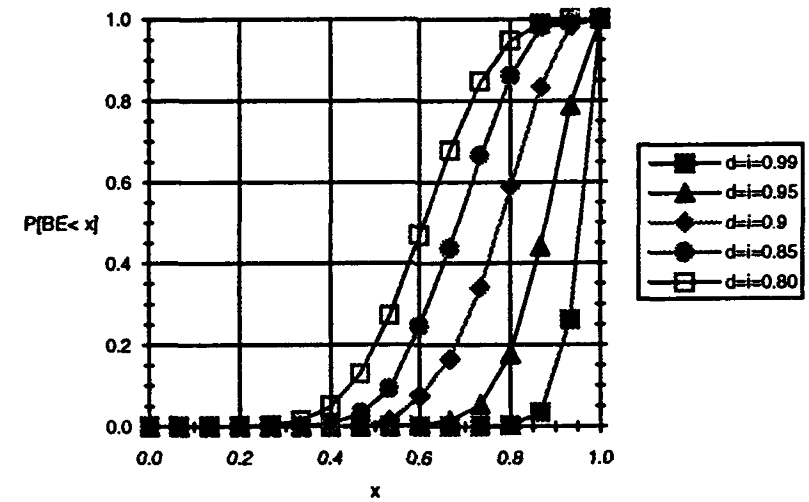


Figure 4.2 BE CDF (Sample Size = 15)

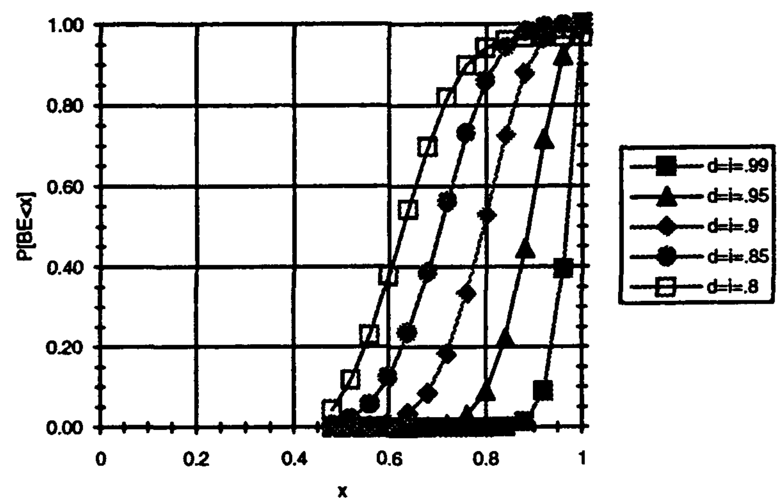


Figure 4.3 BE CDF (Sample Size = 25)

Operational Availability

While A_O is a function of MTBCF and MDT, it is usually dominated by the larger MTBCF value for highly reliable systems -- the result is that the distribution of A_O will be heavily skewed towards 1.0. For this analysis, MDT is assumed constant at its required level of 2.25 hours for any test length and number of failures. In order to analyze the effect of test length on A_O, the range and mean of A_O are observed over the possible number of failures (Poisson distributed). The failure rate is analyzed at three levels:

- $\lambda_1 = 0.0001$ (MTBCF = 10,000 hours; *bench reliability*)
- $\lambda_2 = 0.001$ (MTBCF = 1,000 hours; *intermediate value*)
- $\lambda_3 = 0.0236$ (MTBCF = 42.3 hours; *test requirement*)

The approach used to predict the BE for several sample sizes is also used to predict A_O for a range of test lengths. For λ_1 and λ_2, the number of failures is likely to be small so a table is used to show the results. For larger numbers of failures, such as for λ_3, a graph is more appropriate to show the results.

When λ_1 is assumed, then the probability more than 1 failure is approximately zero (Table 4.1(a)). While the number of possible failures increases for λ_2, the probability of A_O less than 0.95 is still approximately zero for the test lengths observed (Table 4.1(b)). If it is suspected MTBCF≈42 hours, then λ_3 is used and the situation is similar to the M-Demo when BE was predicted to be 0.81 -- increasing test length will decrease the variance of the test results, but it will not significantly increase the probability of passing the test. Figure 4.4 shows that for λ_3:

$P[A_O \geq 0.95] \approx 0.65$ when test length is 50 hours,
$P[A_O \geq 0.95] \approx 0.68$ when test length is 200 hours, and
$P[A_O \geq 0.95] \approx 0.70$ when test length is 350 hours.

Table 4.1 Tabular A_0 PDF

(a) $\lambda_1 = 0.0001$

	50 test hours		200 test hours		350 test hours	
Failures	x	$P[A_0=x]$	x	$P[A_0=x]$	x	$P[A_0=x]$
0	1	.995	1	.980	1	.966
1	.957	.005	.989	.020	.994	.034

(b) $\lambda_2 = 0.001$

	50 test hours		200 test hours		350 test hours	
Failures	x	$P[A_0 \leq x]$	x	$P[A_0 \leq x]$	x	$P[A_0 \leq x]$
0	1	.951	1	.819	1	.705
1	.957	.048	.989	.164	.994	.247
2	.917	.001	.978	.016	.987	.043
3	.881	.000	.967	.001	.981	.005

Accounting for MTBCF in the Test Length Decision. Since MTBCF was a critical measure for the AN/ALR-69 RWR in determining test length, it is necessary to show that it can be accounted for in these calculations (ref RWRLAR:2). Equation (3-9) can be rewritten to calculate MTBCF as a function of A_0 and MDT. Through Equation (4-2), the impact of MDT and A_0 on MTBCF can be studied.

$$MTBCF = \frac{(MDT)(A_o)}{1-A_o} \tag{4-2}$$

If MDT is assumed constant at its requirement of 2.25 hours, Table 4.2 shows MTBCF meets its requirement of 42.3 hours when $A_0 > 0.95$. As MDT increases from 2.25, the lower bound A_0 must attain for MTBCF to meet its requirement increases. Similarly, this lower bound decreases as MDT decreases from 2.25.

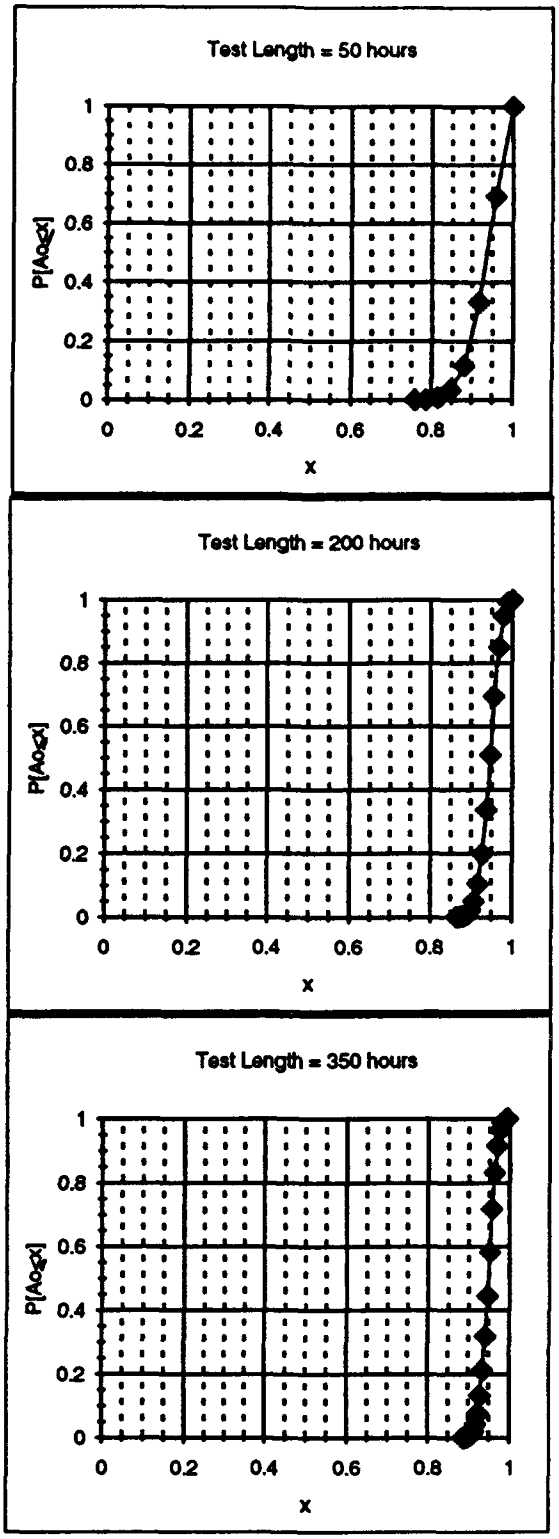


Figure 4.4 A_0 CDF: MTBCF=42 hours

Table 4.2 A_o and MTBCF Relationship

A_o	1.0	0.975	0.95	0.925	0.90
MTBCF (hours)	∞	87.75	42.75	27.75	20.25

Operational Suitability

For this analysis, the multiplicative function will be used to illustrate predicting OS. Whether a multiplicative or additive utility function is used, OS will be a function of A_O, BE, and the weighting factor (α). Once values for A_O and BE are obtained, the value of α is used to determine the value of OS and its test requirement. As it turns out, the CD function and additive function produce identical OS values for any α when $BE=A_O$. As the BE and A_O values diverge, the curvature of the CD indifference line becomes more noticable. This point is best shown by approximating the largest likely difference between BE and A_O. If this difference is 0.4, which should be a conservative estimate for highly reliable systems, then Figure 4.5 shows that the CD function and the additive function produce similar results. A factor that makes the CD function more desirable than the additive function for this analysis is that it produces a slightly more conservative OS value (OS(mult)≤OS(add)).

Multiplicative Utility Function. For the OS calculations, Table 4.3 shows the calculations used to predict OS for three cases. For each case, sample size of 25 and test length of 200 are used to calculate the interval containing OS values is observed. Figure 4.6 plots BE and A_O versus α -- OS is determined by choosing α to weight BE and A_O. The appropriate α will vary depending on the particular system being tested. Parts (a), (b), and (c) of Table 4.3 and Figure 4.6 correspond directly.

By not specifying α, the effect of the weight choice on the OS measure can be studied. Figure 4.6(a) shows that the OS lower bound for the interval meets the test requirement for all values of α. Figure 4.6(b) shows that the lower bound meets the test requirement for $\alpha \leq 0.3$. Part (c) shows that the lower bound does not meet the requirement for any α.

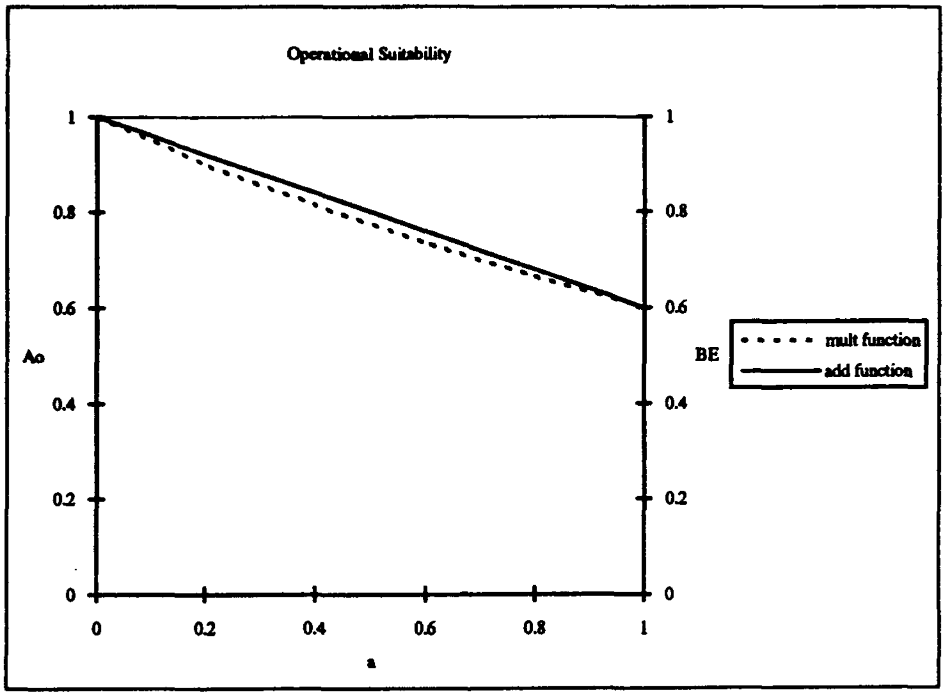


Figure 4.5 Comparison of Additive and Multiplicative Utility Functions

Calculating BE, A_O, and Operational Suitability from Test Results

The functions developed in Chapter III can be used with test results to calculate BE, A_O, and OS. The quantified suitability measure could aid in determining whether the system is operationally suitable. The actual test results for the AN/ALR-69 RWR are an example of perfect performance. Hypothetical results of lesser performance are shown for completeness. Table 4.4 shows operational suitability OS and the test results used to calculate it. The range of OS reflects the choice of α to weight BE and A_O. Figure 4.7 shows the range of OS for each of these cases as α varies from 0 to 1.

Table 4.3 Calculation of OS Range and Mean (99% BE and A_o Intervals)

(a)

BE: d=i=.99			
sample size	5	15	25
99% Interval	(0.8, 1.0)	(0.87, 1.0)	**(0.92, 1.0)**
90% Interval	(1.0, 1.0)	(0.93, 1.0)	(0.96, 1.0)
80% Interval	(1.0, 1.0)	(0.93, 1.0)	(0.96, 1.0)
mean BE	0.980	0.980	**0.980**

$A_o: \lambda_1=0.0001$			
test length	50	200	350
99% Interval	(0.957, 1.0)	**(0.989, 1.0)**	(0.994, 1.0)
90% Interval	(1.0, 1.0)	(1.0, 1.0)	(1.0, 1.0)
80% Interval	(1.0, 1.0)	(1.0, 1.0)	(1.0, 1.0)
mean A_o	0.999	**0.999**	0.999

Suitability						
α=	0	0.2	0.4	0.6	0.8	1
High	1	1	1	1	1	1
Mean	0.999	0.996	0.992	0.988	0.984	0.980
Low	0.989	0.975	0.961	0.947	0.933	0.920
Test Req	0.95	0.920	0.891	0.863	0.836	0.81

(b)

BE: d=i=.95			
sample size	5	15	25
99% Interval	(0.6, 1.0)	(0.67, 1.0)	**(0.76, 1.0)**
90% Interval	(0.8, 1.0)	(0.80, 1.0)	(0.84, 1.0)
80% Interval	(0.8, 1.0)	(0.87, 1.0)	(0.88, 1.0)
mean BE	0.903	0.903	**0.903**

$A_o: \lambda_2=0.001$			
test length	50	200	350
99% Interval	(0.957, 1.0)	**(0.978, 1.0)**	(0.994, 1.0)
90% Interval	(1.0, 1.0)	(0.989, 1.0)	(0.994, 1.0)
80% Interval	(1.0, 1.0)	(1.0, 1.0)	(1.0, 1.0)
mean A_o	0.998	**0.998**	0.998

Suitability						
α=	0	0.2	0.4	0.6	0.8	1
High	1	1	1	1	1	1
Mean	0.998	0.978	0.959	0.939	0.921	0.903
Low	0.978	0.930	0.884	0.841	0.799	0.760
Test Req	0.95	0.92	0.891	0.863	0.836	0.81

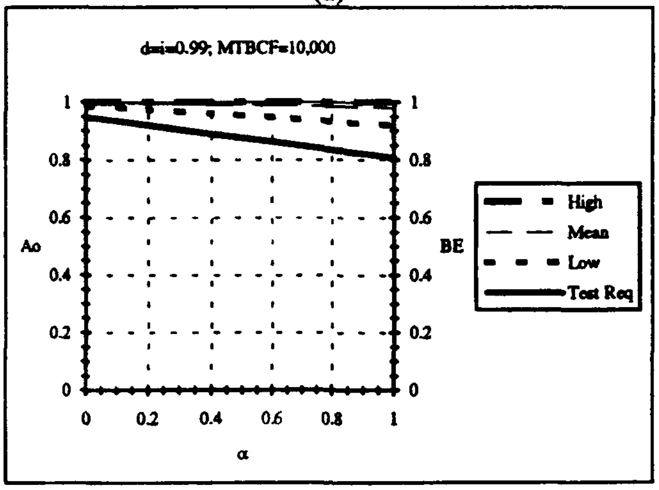
(c)

BE: d=i=.90			
sample size	5	15	25
99% Interval	(0.4, 1.0)	(0.53, 1.0)	**(0.64, 1.0)**
90% Interval	(0.6, 1.0)	(0.67, 1.0)	(0.72, 1.0)
80% Interval	(0.6, 1.0)	(0.73, 1.0)	(0.76, 1.0)
mean BE	0.81	0.81	**0.81**

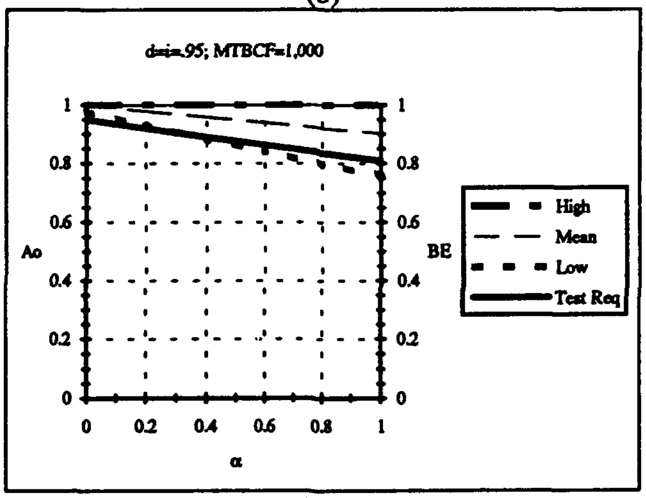
$A_o: \lambda_2=0.0236$			
test length	50	200	350
99% Interval	(0.847, 1.0)	**(0.899, 1.0)**	(0.907, 1.0)
90% Interval	(0.881, 1.0)	(0.927, 1.0)	(0.928, 1.0)
80% Interval	(0.917, 1.0)	(0.937, 1.0)	(0.934, 1.0)
mean A_o	0.949	0.949	0.949

Suitability						
α=	0	0.2	0.4	0.6	0.8	1
High	0.991	0.993	0.995	0.996	0.998	1
Mean	0.950	0.920	0.891	0.863	0.836	0.810
Low	0.899	0.840	0.785	0.733	0.685	0.640
Test Req	0.95	0.920	0.891	0.863	0.836	0.81

(a)



(b)



(c)

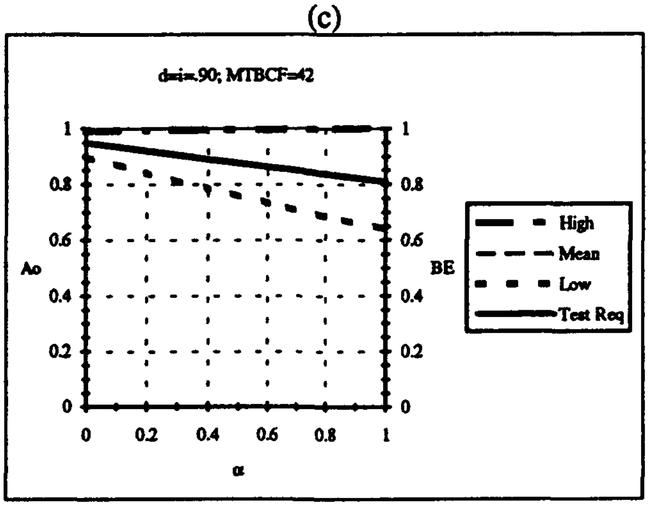


Figure 4.6 OS Range, Mean, and Test Requirement (99% BE and A_0 Intervals)

Table 4.4 Calculating OS (multiplicative) from Test Results

	Test Results			
	Actual	Test #1	Test #2	Test Req
d=FDR	1.00	0.95	0.95	0.90
i=FIR	1.00	0.95	0.80	0.90
BE (M-Demo)	1.00	0.90	0.76	0.81
MTBCF	∞	21 hours	110 hours	42 hours
MDT	0 hours	2.25 hours	2.25 hours	2.25 hours
A_o	1.00	0.90	0.98	0.95
OS	1.00 ≤ OS ≤ 1.00	0.90 ≤ OS ≤ 0.90	0.76 ≤ OS ≤ 0.98	0.81≤ OS ≤ 0.95

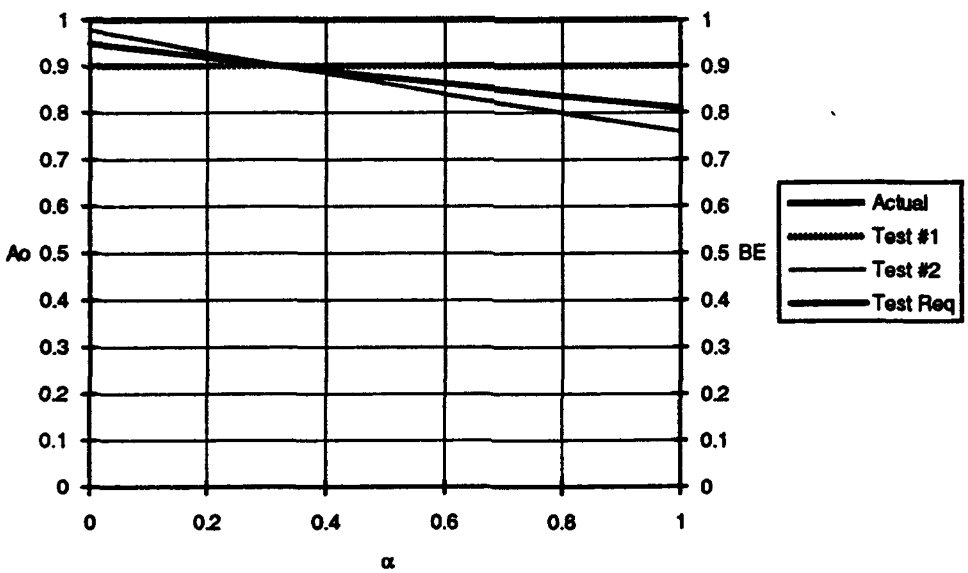


Figure 4.7 OS Calculated from Test Results

The following conclusions can be drawn from Figure 4.7:

- Test #1 results are more desirable than Test #2 results and meet test requirements when $\alpha \geq 0.3$.

- Test # 2 results are more desirable than Test #1 results and meet test requirements when $\alpha < 0.3$.

Summary

In this chapter, the measures for BE, A_O, and OS were analyzed for various levels of system and BIT performance. It was shown that how these measures could be used to determine the appropriate test length and M-Demo size. Only an interval for OS is provided in each case. There will be a OS distribution for each value of α, but this derivation is deferred to future research. On a basic level, OS values can be used to rank order different test outcomes. However, in order to answer 'how much better is OS=.95 than OS=.90?,' factors such as test costs and the cost of poor performance should be included in a more sophisticated OS function. Further study in this and other areas is proposed in Chapter V.

V. Conclusions

Summary

In review, the current procedures being used to assess OS during IOT&E do not result in a quantitative measurement of OS that could an be used to assess the value of additional testing (increased test length or sample size). This research then showed that such a measure could be calculated as a function of BE and A_O and potentially used to assess the value of additional testing (increased test length or sample size). Of the two approaches presented to quantify BE during a time-terminated test, the second approach was preferred because it could be used with a M-Demo. Using A_O to determine the flight/ground test length was justified by showing its relationship to the critical measure, MTBCF.

Probability statements were made about BE and A_O from the prior information, but this analysis revealed the reliance of the results on the accuracy of the prior information -- the results are only as reliable as the prior information. In calculating OS, it was shown that the multiplicative function produces measurements similar to the additive function, but is preferred since it produces a more conservative result.

The objective of this research, which was to provide a method to assess the value of testing a system in support of the full-rate production decision, was partially met in that a quantitative measure for OS was developed. As mentioned in Chapter IV, the next step is to enhance the OS utility function with factors such as test costs.

Recommendations

As this research was performed, it was determined that more research is required in several areas.

Other Operational Suitability Applications. The methodology developed in this chapter must be tailored before it can be applied to another system. It would be useful to determine what commonalities exist among the many types of systems tested by AFOTEC. Discussions with personnel in AFOTEC/SAL revealed that while the general test philosophy is common for most systems, details such as test measures and their interdependence are system specific.

System Effectiveness. Similarly, the ideas developed in this research could be applied to system effectiveness. This application is also challenging because the relationships between system effectiveness measures tend to be system specific.

Ideal T&E Value Function. The functions developed in this research do not address costs or system effectiveness. In addition to system effectiveness, it should be possible to integrate costs (such as test costs, production and operation and supply (O&S) costs) into the test length/sample size decision.

The form of the suitability function (e.g., Cobb-Douglas) should be researched to determine the effect of the function form on results. Regardless of the function form, the ideal T&E value function would include at least quantitative measurements of system effectiveness, operational suitability and costs. This value function could be used for Bayesian Analysis.

Standardizing BE for M-Demo and Continuous-time Test. The current M-Demo approach does not use the actual distribution of BIT-detectable failures. Research into the representation of this distribution could improve the BE measure.

Continuous-time Markov Chain Model. It was suggested during this research that the flight/ground test could be modeled as a continuous-time Markov chain. This idea presents the ability to not only model the performance of the system, but to model the entire operational/repair cycle of the system.

Bibliography

1. Clemen, R. T. Making Hard Decisions: An Introduction to Decision Analysis, PWS-Kent Publishing Company, 1991.

2. Conlon, Dr. John C., Walter A. Lilius, and Lt. Col. Frank H. Tubbesing, Jr., USAF. Test and Evaluation of System Reliability, Availability, and Maintainability: A Primer. Office of the Director, Defense Test and Evaluation, Under Secretary of Defense for Research and Engineering. Washington, D.C.: 1982.

3. D'Agostino, Captain Paul. Logistics Analysis Review 4 for AN/ALR-69 Radar Warning Receiver. Air Force Operational Test and Evaluation Center, Kirtland AFB, New Mexico.

4. D'Agostino, Captain Paul. Logistics Analysis Review 2 for Military Microwave Landing System Avionics. Air Force Operational Test and Evaluation Center, Kirtland AFB, New Mexico.

5. D'Agostino, Captain Paul. Air Force Operational Test and Evaluation Center, Kirtland AFB, New Mexico. Telephone interview. 18 November 1993

6. D'Agostino, Captain Paul. Reliability Test Guide for Exponential Systems. Air Force Operational Test and Evaluation Center, Kirtland AFB, New Mexico.

7. Department of the Air Force. Test and Evaluation. AF Regulation 80-14 (Draft). HQ USAF: Washington, 11 November 1991.

8. Department of the Air Force. Management of Operational Test and Evaluation. AFOTEC Instruction 99-101. HQ AFOTEC: Kirtland AFB NM, 1 October 1993.

9. Department of the Air Force. Operational Suitability Test and Evaluation. AFOTEC Pamphlet 400-1. HQ AFOTEC: Kirtland AFB NM, 15 May 1991.

10. Greenlee, Don. Terms of Reference for a MORS/ITEA Mini-Symposium to Address "How much testing is enough?" (Draft). 6 June 1993.

11. Kapur, K. C. and L. R. Lamberson. Reliability in Engineering Design. John Wiley and Son, Inc. New York: 1977.

12. Keeney, R. L. "Decision Analysis: An Overview," Operations Research, 30: 803-838 (September-October 1982).

13. Nicholson, Walter. Microeconomic Theory: Basic Principles and Extensions. The Dryden Press. Chicago: 1989.

14. Office of the Secretary of Defense. Under Secretary of Defense Acquisition. Defense Acquisition. DoD Instruction 5000.1. Washington: 23 Feb 1991.

15. Office of the Secretary of Defense. Under Secretary of Defense Acquisition. Defense Acquisition Management Policies and Procedures. DoD Instruction 5000.2. Washington: 23 Feb 1991.

16. Pratt, J. W., H. Raiffa, and R. Schlaifer. "The Foundations of Decision Under Uncertainty: An Elementary Exposition," Journal of the American Statistical Association, 59: 353-375 (June 1964).

17. SYS 229, Test and Evaluation Management Coursebook, School of Systems and Logistics, Air Force Institute of Technology, Wright-Patterson AFB OH, June 1993.

VITA

Captain James N. Serpa was born in 1966 in Pittsburgh, Pennsylvania. He spent his childhood in Cumberland, MD and in the Pittsburgh area. Following graduation from Upper St. Clair High School, he accepted an Air Force ROTC scholarship and entered the Pennsylvania State University at State College in 1984. In 1988, he graduated with a Bachelor of Science degree in Applied Mathematics and a reserve commission in the USAF.

In 1989, Captain Serpa completed Undergraduate Space Training at Lowry AFB CO en route to the 2d Satellite Control Squadron (AFSPACECOM) at Falcon AFB CO. In three years, he progressed from Satellite Operations Officer to Standardization Officer at 2d Space Wing. He met his wife Kerrie in 1989 and they married in 1991. Upon graduating from AFIT, Captain Serpa will be assigned to HQ AETC at Randolph AFB TX.

Permanent address: 1133 Galaxy Circle
Pittsburgh, PA
15241

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

[illegible]

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1994	Master's Thesis

4. TITLE AND SUBTITLE
AN ANALYSIS OF OPERATIONAL SUITABILITY FOR TEST AND EVALUATION OF HIGHLY RELIABLE SYSTEMS

5. FUNDING NUMBERS

6. AUTHOR(S)
James N. Serpa, Captain, U.S. Air Force

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
WPAFB, OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GOR/ENS/94M-13

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
HQ AFOTEC/SAL
8500 Gibson Blvd SE
Kirtland AFB, NM 87117-5558

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this research was to develop a quantitative measure of operational suitability (OS) and determine its applicability in making the test length decision prior to Initial Operational Test and Evaluation (IOT&E). The current approach used by the Air Force Operational Test and Evaluation Center (AFOTEC) was presented and used to establish the relationships of the test measures. It was established that OS could be represented by a function of operational availability (A_O) and built-in test effectiveness (BE). BE was defined and measures proposed based on the method of data collection. A proposal for predicting A_O, BE, and OS to determine the proper test length and sample size was analyzed for several examples of prior information. Multiplicative and additive utility functions were proposed as possible ways to calculate OS. It was shown that probability statements could be made about BE, A_O, and OS from the prior information; this analysis revealed the reliance of the results on the prior information

14. SUBJECT TERMS
Test and Evaluation, Operational Suitability, Decision Analysis
Sample Size Determination

15. NUMBER OF PAGES
50

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL