12-2013

A cost model for managing producer and consumer risk in availability demonstration testing.

Randall L. Walker
University of Louisville

A COST MODEL FOR MANAGING PRODUCER AND CONSUMER RISK IN AVAILABILITY DEMONSTRATION TESTING

By

Randall L. Walker
B.S. Mathematics, University of Louisville, 1988
B.S. Engineering Science, University of Louisville, 1990
M. Eng. Industrial Engineering, University of Louisville, 1991

A Dissertation
Submitted to the Faculty of the
J.B. Speed School of Engineering of the University of Louisville
in Partial Fulfillment of the Requirements
for the Professional Degree of

Doctor of Philosophy

Department of Industrial Engineering
University of Louisville
Louisville, Kentucky

December 2013

A COST MODEL FOR MANAGING PRODUCER AND CONSUMER RISK IN AVAILABILITY DEMONSTRATION TESTING

By

Randall L. Walker
B.S. Mathematics, University of Louisville, 1988
B.S. Engineering Science, University of Louisville, 1990
M. Eng. Industrial Engineering, University of Louisville, 1991

A Dissertation Approved On

December 4, 2013

By the following Dissertation Committee:

_____
Dr. John S. Usher, Dissertation Director

_____
Dr. Gail W. DePuy

_____
Dr. Lihui Bai

_____
Dr. C. Tim Hardin

_____
Dr. Anup Kumar

ACKNOWLEDGEMENT

I would like to thank my dissertation director Dr. John Usher for his mentoring, guidance and support throughout this process. As a non-traditional student, I tested his patience to the limit and am very thankful for his resolve. I would like to thank Dr. Gail DePuy for the motivation, guidance and encouragement she has provided over many years. I would like to thank Dr. Lihui Bai for her perspective in pointing out aspects of the research I hadn't considered. I am grateful to Dr. Tim Hardin for his time and assistance with the cost modeling portion of this work. His enthusiasm and genuine care for his students, colleagues, and friends are truly motivating. I would like to thank Dr. Anup Kumar for his time and guidance while evaluating this work.

My mother has always been an inspiration to me; her resilience while confronting life's overwhelming competing objectives is remarkable. I am proud of her and grateful for her lifelong support.

I would also like to thank my family and friends. I have been continually reminded of the significance of a strong network of people whom I can depend on, and who can depend on me. To be part of something that is always bigger than myself provided the perspective I needed to keep this dissertation in perspective.

Most of all, I would like to thank my wife Sharon and daughter Bailye for their unconditional love and support throughout this process. I disrupted your lives to pursue this goal, and would never have finished without your encouragement and confidence.

ABSTRACT

A COST MODEL FOR MANAGING PRODUCER AND CONSUMER RISK IN AVAILABILITY DEMONSTRATION TESTING

Randall L. Walker

December 4, 2013

Evaluation and demonstration of system performance against specified requirements is an essential element of risk reduction during the design, development, and production phases of a product lifecycle. Typical demonstration testing focuses on reliability and maintainability without consideration for availability. A practical reason considers the fact that demonstration testing for availability cannot be performed until very late in the product lifecycle when production representative units become available and system integration is completed. At this point, the requirement to field the system often takes priority over demonstration of availability performance.

Without proper validation testing, the system can be fielded with reduced mission readiness and increased lifecycle cost. The need exists for availability demonstration testing (ADT) with emphasis on managing risk while minimizing the cost to the user. Risk management must ensure a test strategy that adequately considers producer and consumer risk objectives.

This research proposes a methodology for ADT that provides managers and decision makers an improved ability to distinguish between high and low availability systems. A new availability demonstration test methodology is defined that provides a

useful strategy for the consumer to mitigate significant risk without sacrificing the cost of time to field a product or capability. A surface navy electronic system case study supports the practical implementation of this methodology using no more than a simple spreadsheet tool for numerical analysis.

Development of this method required three significant components which add to the existing body of knowledge. The first was a comparative performance assessment of existing ADT strategies to understand if any preferences exist. The next component was the development of an approach for ADT design that effectively considers time constraints on the test duration. The third component was the development of a procedure for an ADT design which provides awareness of risk levels in time-constrained ADT, and offers an evaluation of alternatives to select the best sub-optimal test plan.

Comparison of the different ADT strategies utilized a simulation model to evaluate runs specified by a five-factor, full-factorial design of experiments. Analysis of variance verified that ADT strategies are significantly different with respect to output responses quality of decision and timeliness. Analysis revealed that the fixed number of failure ADT strategy has the lowest deviation from estimated producer and consumer risk, the measure of quality. The sequential ADT strategy had an average error 3.5 times larger and fixed test time strategies displayed error rates 8.5 to 12.7 larger than the best. The fixed test time strategies had superior performance in timeliness, measured by average test duration. The sequential strategy took 24% longer on average, and the fixed number of failure strategy took 2.5 times longer on average than the best.

The research evaluated the application of a time constraint on ADT, and determined an increase in producer and consumer risk levels results when test duration is

limited from its optimal value. It also revealed that substitution of a specified time constraint formatted for a specific test strategy produced a pair of dependent relationships between risk levels and the critical test value. These relationships define alternative test plans and could be analyzed in a cost context to compare and select the low cost alternative test plan. This result led to the specification of a support tool to enable a decision maker to understand changes to α and β resulting from constraint of test duration, and to make decisions based on the true risk exposure. The output of this process is a time-constrained test plan with known producer and consumer risk levels.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1

INTRODUCTION

1.1 Motivation

Common design approaches such as open system architecture and software configurable functionality have led to improved integration of technologies and information management. This has enabled system designers to synthesize physical configurations with ever increasing complexity and functionality. The requirements for a system to be both available for use when needed and to be cost effective over the lifecycle are critical for users of large-scale, complex end items such as the Department of Defense (DoD). These concurrent requirements are critical for major defense acquisition programs (MDAP) managed by the DoD. As such, program managers are directed to include reliability, maintainability, availability, (RMA) and lifecycle cost requirements within the specification of critical system characteristics (Joint Requirements Oversight Council, 2012). Quantification of these critical characteristics is designed to be numerical in nature so they can be described by a mathematical equation and evaluated throughout the system's life, from concept to disposal.

In March 2011, the United States Government Accountability Office (GAO) reported to congressional committees on the status of the DoD 2010 major program portfolio. GAO stated that of the 98 programs in the portfolio, with total lifecycle budgetary obligations of $1.68 trillion, nearly 50% were failing to meet cost performance

goals. Over $70 billion of cost growth seen in these programs could not be accounted for due to changes in contracted quantities (Government Accountability Office, 2011). The 2013 update of this report stated similar findings. The 86 programs in the 2012 portfolio have experienced nearly $400 billion in cost growth from initial program cost estimates. In fact, cost growth represented forty-one percent of the funding needed to complete these programs (Government Accountability Office, 2013).

The annual report to Congress by the Director, Operational Test and Evaluation (DOT&E) summarizes results of initial operational test and evaluation reports on DoD systems. In the 2012 report, DOT&E stated that only 57% of systems tested from 1997 through 2012 met minimum reliability requirements. In fact, only 53% of systems evaluated during 2012 met minimum requirements (Director, Operational Test and Evaluation, 2012). These tests focus on the components of reliability and maintainability, but an assessment of availability is not typically made.

The references above are not intended to criticize the performance of program managers and their dedicated government/industry teams. Instead, it highlights the difficulty in achieving availability and cost goals in a complex and challenging environment. The competing objectives of maximum availability with minimum life cycle cost must be balanced within a comprehensive system engineering process. Many design techniques can improve system availability including redundancy, channelized graceful degradation, and use of high reliability components. However, these methods result in increased acquisition and operational support costs over the lifecycle of a system. This challenge is further amplified by the need to balance cost and availability

across multiple product baselines and configurations for end items that utilize spiral development cycles such as long life systems within the DoD.

For large scale systems, overarching requirements are allocated and specified at sub-system and component levels during functional analysis and allocation within the systems engineering process. This is depicted in the decomposition half of the systems engineering 'V' shown in Figure 1. Design trade-offs and configuration changes are evaluated on a piece of the system, often without total definition of the performance impact to the system as a whole. As product realization begins, design synthesis and developmental testing is done on sub-systems that are later brought together for integration. A comprehensive understanding of intersystem dependency and interface performance is not known until the final product is integrated and operated in a realistic environment. Regardless of the knowledge gained from developmental testing, or modeling and simulation performed from conception through the production cycle, the ultimate performance of the complete system can only be assessed through demonstration testing of the fully integrated end product.

Standard reliability engineering textbooks document the recommended guidelines and processes to perform demonstration testing for the reliability and maintainability characteristics of a system. However, demonstration testing for system availability is routinely omitted. Even though availability is a function of system reliability and maintainability, the intersystem dependencies and interface issues discussed above cannot be arbitrarily dismissed. Furthermore, in the face of imperfect maintenance, reliability demonstration testing may not adequately quantify the system time to failure distribution. The ability to perform availability demonstration testing (ADT) is essential to evaluate a

complex system's availability characteristics, and determine the impact to life cycle cost. This test methodology must be robust in the face of schedule and budgetary pressures realized by DoD weapon system program managers.

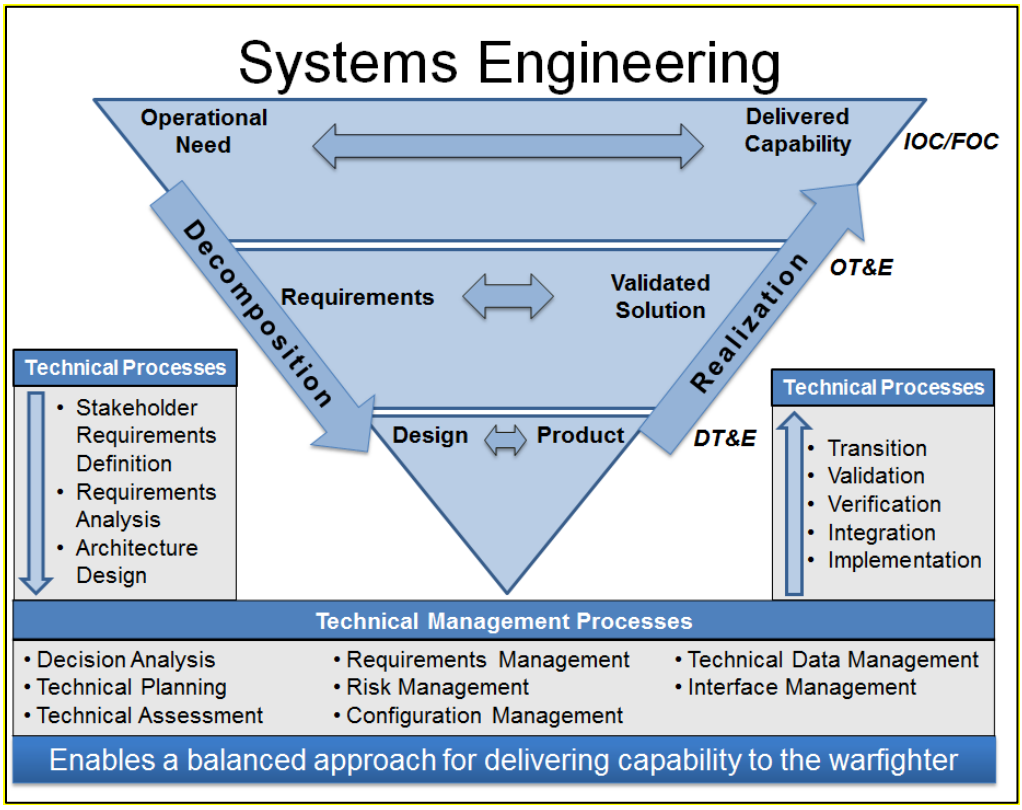

Figure 1. Systems Engineering Vee Diagram (source: The Defense Acquisition Guidebook, https://dag.dau.mil/Pages/Default.aspx, accessed on 5 Dec 2013)

1.2 Problem Statement

Evaluation and demonstration of system performance against specified requirements is an essential element of risk reduction during the design, development, and production phases of a product lifecycle. Typical demonstration testing focuses on reliability and maintainability without consideration for availability. In some paradigms, it is considered unnecessary due to the dependency between the three elements of RMA that is well defined in closed form equations. For example, MIL-STD-961 states that a defense standard with quantitative requirements for both reliability and maintainability

should not state an availability requirement (Department of Defense, 2008). A more practical reason considers the fact that demonstration testing for availability cannot be performed until very late in the product lifecycle when production representative units become available and system integration is completed. At this point, the requirement to field the system often takes priority over demonstration of availability performance.

The purpose of demonstration testing is to statistically validate that a system meets design criteria. Without the risk reduction provided by this type of testing, a consumer can take delivery of a system that fails to meet a critical performance characteristic. When this performance is availability, the outcome is reduced mission readiness and increased lifecycle cost. Decreased availability due to a higher failure rate increases the probability that a system is not operational when called upon, or fails before a mission is complete. Unavailability due to maintainability or sustainment issues will result in a higher number of repair actions with an associated increase in operational support cost. In its worst case, maintainability issues can result in a larger logistics tail with significant cost impact.

A need exists for availability demonstration testing with emphasis on managing risk while minimizing the cost to the user. Risk management must ensure a test strategy that adequately considers producer and consumer risk objectives. Cost should consider not only the fiscal aspect, but also the cost impact of time to market or fielding of the system.

The purpose of this research was to investigate and propose a methodology for availability demonstration testing (ADT) that provides users of complex, large scale systems an improved ability to distinguish between high and low availability systems.

This methodology recognizes the competing objectives of high availability and low life cycle cost, and focuses consideration for test cost in terms of fiscal aspects. Four objectives were established to achieve this desired outcome:

- The first objective of this research was to evaluate and document the current state of the art in ADT, across both the public and private sectors.
- The second objective was the evaluation of ADT methodologies to assess comparative performance and understand if any preferences exist.
- The third objective was to develop an approach for ADT design that recognized time constraints of a decision maker or test manager.
- The fourth was the definition of a procedure for ADT design which implements all results of the research into a methodology enabled by a basic spreadsheet tool for implementation.

The first objective involved a literature review on specific topics within the broad scope of RMA. These included general definitions for system availability, the role of test and evaluation within the systems engineering process, and current research in ADT. These results establish the current body of knowledge and existing challenges.

The second objective utilized simulation modeling with Arena® software to perform a design of experiments on ADT. Factors under consideration included producer and consumer risk levels, availability required under the producer risk, discrimination ratio, the mean of the time to failure distribution, and the test methodology. The goal of this effort was to determine the statistical significance of test methodologies identified in the literature review with respect to quality and timeliness.

The third objective was focused on creation of a methodology to define a desirable ADT strategy in the face of time constraints. This method utilized an equation for risk mitigation cost to define a balance between producer and consumer risk that was acceptable to a decision maker. A simple Excel™ spreadsheet was created to provide the

numerical tool for executing this approach. For the purpose of this research, cost consideration focuses on the fiscal attribute and does not consider the opportunity cost of time.

The fourth objective was the creation of a procedure to implement the overall outcomes of this research into a well-defined approach for ADT under time constraints. This approach included a process flow diagram with definition of data inputs, analytical methods, and outputs for each step. This process was demonstrated with a case study.

1.3 Research Significance

The development of complex, large scale systems is required to meet the needs of DoD end users. The maturation and integration of new technology, as well as the implementation of existing technology in new applications, generates technical risk within these development efforts. Demonstration testing is a primary means of mitigating this risk and assuring the consumer that a costly investment will meet stated requirements. While reliability and maintainability demonstration testing is well known and implemented, ADT is less known and rarely implemented. The consumer's desire to field the end item, coupled with the need of the producer to receive timely payment, creates pressure to skip the risk reduction step of ADT. This exposes the consumer to the potential of reduced mission success and increased operational support costs. This research defines an ADT methodology that reduces the time to make a decision by providing a useful strategy for the consumer to mitigate significant risk without sacrificing the cost of time to field a product or capability.

1.4 Dissertation Organization

This dissertation is written in manuscript format and is organized as follows. In Chapter 2, a literature review is presented, including literature pertaining to basic definitions and nomenclature for availability, the systems engineering process with attention to test and evaluation within the control loop, ADT methodologies, and test and evaluation cost models. Chapter 3 presents a comparative evaluation of alternative ADT methodologies with performance measures of quality and timeliness. Chapter 4 focuses on the development of a methodology for defining an ADT strategy in the presence of a time constraint. Chapter 5 develops and presents an implementation process to be used by a decision maker or test developer that incorporates the methodology established in prior chapters. Lastly, Chapter 6 includes discussion on conclusions, contributions to the body of knowledge, and areas for future research.

CHAPTER 2

LITERATURE REVIEW

This literature review divides the relevant literature into four categories: availability definitions and equations, risk management and mitigation within the systems engineering control loop, literature related to ADT, and literature related to cost modeling for test and evaluation.

2.1 Availability

Availability is defined as "the probability that a system or component is performing its required function at a given point in time or over a stated period of time when operated and maintained in a prescribed manner" (Ebeling, 2010). It is a function of the reliability, maintainability and supply support attributes of a system, and translates these characteristics into a measure of effectiveness for system performance. It is essentially a figure of merit that predicts the system's operational state when called upon for service. Availability and its components are referenced by many acronyms and abbreviations. This research uses exclusively the term RMA to mean reliability, maintainability, and availability.

Availability is the ratio of a system's uptime to total time. This research focuses specifically on repairable systems with binary states, either functional or non-functional. Throughout the study, the variable *x* will represent a concurrent period of operation (time to failure), while *y* will represent a concurrent period of downtime (time to restore, reset,

repair, or replace). Table 1 provides a summary of different representations of the reliability characteristic of a system, or the measures of uptime (Department of Defense, 2005). This description begins with the most basic measure and adds complexity up to the most complex parameter from an analytical perspective.

Table 1

Quantitative Measures of Reliability

Parameter	Description
Mean Time to Failure (MTTF)	Basic measure of reliability for non-repairable items or systems. Average failure-free operating time, during a particular measurement period under stated conditions.
Mean Time Between Failure (MTBF)	Basic measure of reliability for a repairable items or system. The average time during which all parts of the item perform within their specified limit, during a particular measurement period under stated conditions. (NOTE: Does not include system downing events due to preventive maintenance.)
Mean Time Between Maintenance (MTBM)	Basic measure of reliability for repairable fielded systems. The average time between all system maintenance actions. Maintenance actions may be for repair or preventive purposes.
Mean Time Between Repair (MTBR)	Basic measure of reliability for repairable fielded systems. The average time between all system maintenance actions requiring removal and replacement or in-situ repairs of a box or subsystem.
Mean Time Between Critical Failure (MTBCF)	Measure of system reliability that includes the effects of any fault tolerance that may exist. The average time between failures that cause a loss of a system function defined as 'critical' by the consumer.
Mean Time Between Operational Mission Failure (MTBOMF)	Measure of operational mission reliability for the system. The average time between operational mission failures which cause a loss of the system's 'mission' as defined by the consumer.

The reliability parameters indicate the two ways in which a system can move from a functional state to a non-functional state. First, occurrence of an unexpected system malfunction dictates corrective maintenance. Otherwise, a scheduled maintenance action occurs, known as preventive maintenance.

Table 2 provides a summary of different representations of the maintainability characteristic of a system, or the basic measures of downtime (Department of Defense,

2005). As before, this description begins with the most basic measure and adds complexity up to the most complex parameter from an analytical perspective.

Table 2

Quantitative Measures of Maintainability

Parameter	Description
Mean time to repair (MTTR)	Mean maintenance cycle time for corrective maintenance assuming an ideal environment.
Mean Active Maintenance Time (MAMT)	Mean maintenance cycle time for all maintenance (corrective and preventative) excluding logistic and administrative delays.
Mean Down Time (MDT)	Mean system down time including all scheduled and corrective maintenance as well as logistics delays (personnel, material, equipment), and administrative delays (authorization, planning, documentation, approval)

There are generally three types of system availability considered in the design, development, production, and fielding of a complex system. These types are inherent availability (A_i), achieved availability (A_a), and operational availability (A_o) (Ebeling, 2010). All three will be defined, but it is noted that the DoD only references A_i and A_o (Department of Defense, 2008).

Inherent availability, A_i, is a design parameter that assumes a perfect, ideal maintenance environment without scheduled (preventative) maintenance and is defined as:

$$A_i = \frac{MTBF}{MTBF+MTTR} \quad (1)$$

Achieved availability, A_a, includes both corrective and scheduled maintenance. It is frequently used during developmental and production testing where the support environment is not established. The form is defined as:

$$A_a = \frac{MTBM}{MTBM+MAMT} \quad (2)$$

Operational availability, A_o, measures performance in the most realistic circumstances referred to as an operational environment. A_o is defined as:

$$A_o = \frac{MTBM}{MTBM+MDT} \quad (3)$$

In the case of A_o, the measure of uptime can be any of the lower three in Table 1. Use of the MTBM measure assumes that all failures result in loss of system operation. In this case, all corrective maintenance is included in assessment of the metric. If the system has redundancy or graceful degradation, then some corrective maintenance will be applied to repair of items for which the failure does not create a loss of operations. In this case, MTBCF is the uptime measure of choice. Finally, MTBOMF allows for the situation in which a critical failure occurs, but is corrected and operations restored without loss of the mission. The consumer has responsibility to know which measure is most applicable for their intended use, and ensure that the definition of A_o contains the proper term for definition of uptime.

Evaluation of system availability can use any of the three forms presented. The specific objectives of the analysis will dictate which form, and what data to utilize. Early in the design process, inherent availability is used in the allocation of system budget across the sub-systems being developed. As the design takes form, maintenance engineering products will begin to form and allow for a migration to achieved availability. As the system transitions through production and into fielding, operational availability takes on importance as the relevant measure.

Testing to evaluate or demonstrate system performance can only use the form of availability for which data is available. It is also important to validate all parametric assumptions in testing. Parametric distributions are useful in modeling both uptime and downtime values. It is important to understand the characteristics of performance when selecting an appropriate distribution. For instance, a system subject to random failure is likely to be best modeled with an exponential distribution, while a system with wear out failure modes is better modeled with a Weibull or Lognormal distribution (Kelton, Sadowski, and Swets, 2010). Likewise, while it might be reasonable to assume a specific distribution for the mean time to repair, it may not be reasonable to assume the same distribution for the mean down time which also includes logistic and administrative delays. Therefore, the form of availability chosen for test and evaluation should consider both data readiness as well as the appropriateness of assumptions on parametric distributions used within a model.

2.2 System Engineering

The successful development of a dependable large scale or complex system requires the discipline of the systems engineering process. This process translates user requirements into a physical system baseline that meets stated needs and can be produced, fielded, and sustained for the expected life of the product. Requirements can be specified for system functionality, performance level, or to define constraints. Functional requirements define tasks, actions, or activity that must be provided by the system. Performance requirements state the level or extent to which the system must meet a function or set of functions. Constraints are restrictions representative of legal,

political, procedural, moral, technological, and interface conditions [Department of Defense, 2005]. Availability is a performance requirement of the system.

2.2.1 Systems Engineering Process Definition

Figure 2 provides a graphical representation of the systems engineering process. High level requirements are input into a recursive process that decomposes these needs into first a functional, then physical baseline. Throughout this process, competing objectives are identified and managed in order to ensure the end product satisfies all of the users' expectations.

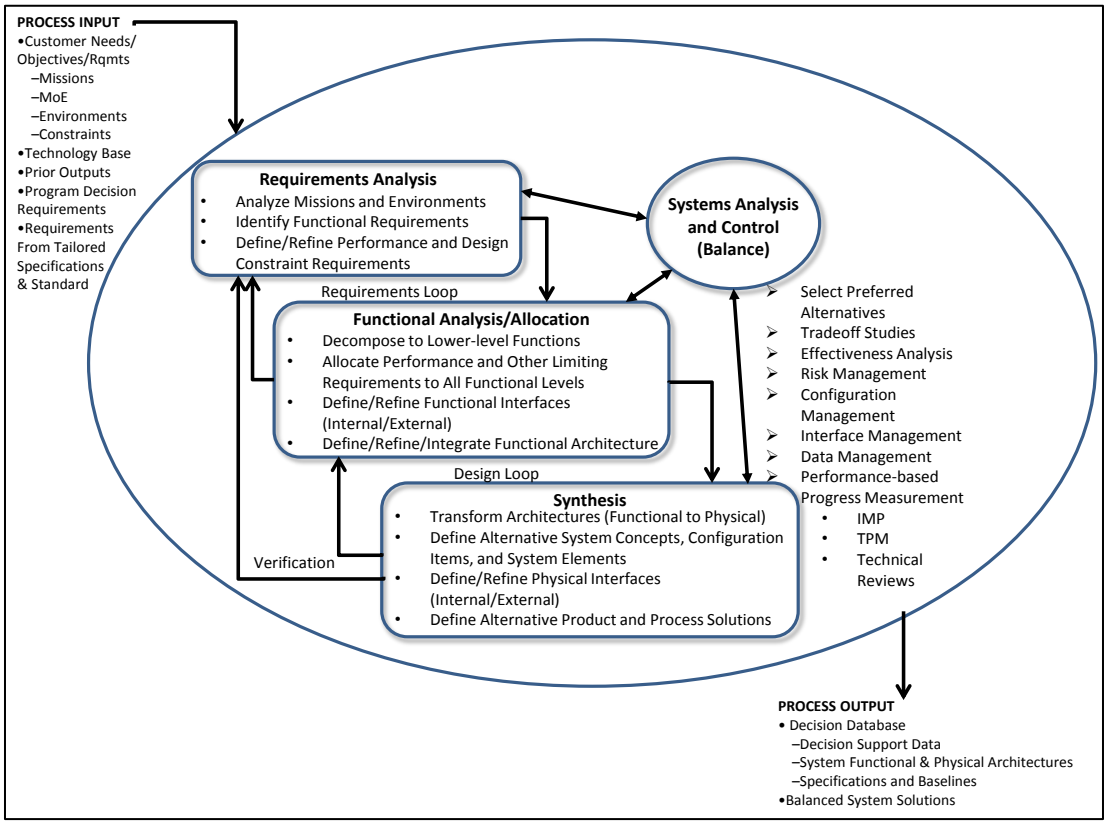

Figure 2. Systems Engineering Process Diagram [source: Department of Defense, 2005]

Three management processes exist within the design bubble. They include the requirements loop, the design loop, and the control loop. Availability requirements are defined and documented within the requirements loop. These are then allocated across

the physical baseline within the design loop. Consideration for availability requirements inside these loops is essential to ensure that performance is adequately specified and that environmental conditions are documented.

Consider the system diagram shown in Figure 3. The system is composed of building blocks that represent the product hierarchy and the life cycle processes required to design, produce, field, operate, support, and dispose of the product. As an example, a surface combatant system such as a Navy destroyer might have a ship system product and a combat system product.

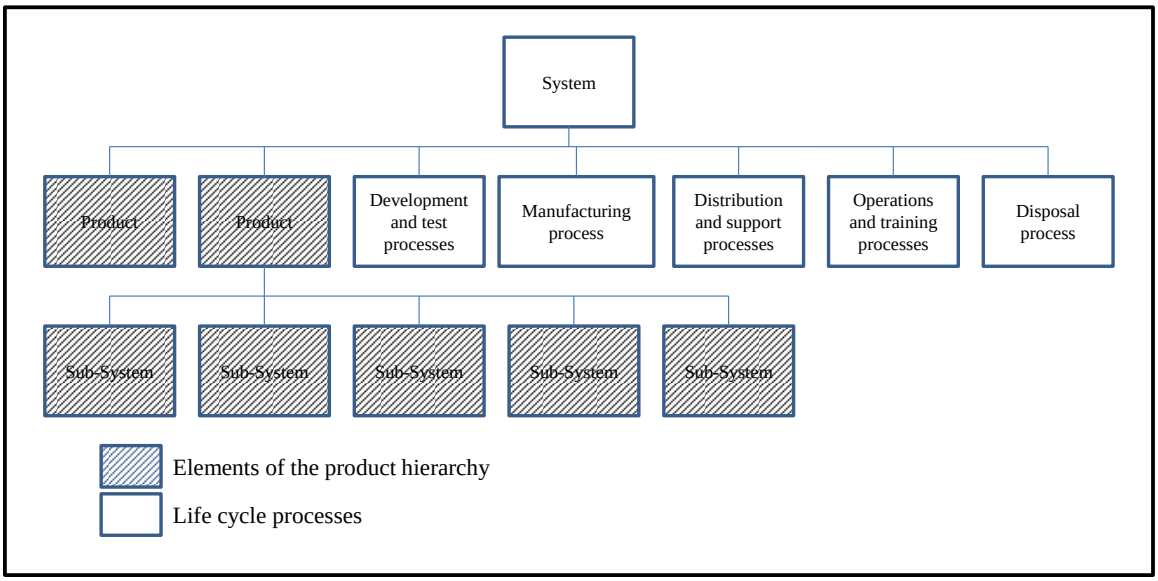

Figure 3. Basic Building Blocks of a System [source: IEEE, 2011]

The combat system product might have an air warfare subsystem, a surface warfare subsystem, an undersea warfare subsystem, and a support subsystem. In a real world example, the hierarchy would be developed to six or seven layers deep. Availability requirements are defined at the system level, and allocated down through the product hierarchy to its lowest level. Many of these product requirements are dependent on life cycle processes within the system and link to critical process requirements. For

example, the operations and training processes and the distribution and support processes will impact the operations of the system, as well as define the logistics processes required to restore functionality in the event of a system failure.

During design and development, these allocated budgets and dependent process requirements must be managed to ensure the resulting system operates in accordance with the users' expectations. As the physical design is matured, each layer of the hierarchy represents the integration of lower level elements and an increase in hardware, software, and operational complexity. Dependent requirements must be validated and reconciled. Failure to fulfill any of these requirements at the system level could result in operational readiness or lifecycle cost breaches. This highlights the importance of demonstrating system availability prior to acceptance of an expensive, long life system.

The control loop within the system engineering process is of significant interest. Within this process, effectiveness analysis and trade-off studies are used to evaluate competing objectives and manage allocated budgets. Furthermore, the risk management process is linked tightly with test and evaluation to ensure that critical performance requirements are demonstrated prior to acceptance of the end item by the customer. Availability requirements are directly linked to mission success rates and life cycle cost, both of which are critical performance parameters. It is important to note that these two critical attributes are often in competition since design approaches which increase mission reliability will often have a negative impact on acquisition cost as well as operation and support cost. There is significant management associated with balancing these competing requirements while coordinating, tracking, and validating the trade-off studies performed simultaneously during design and development.

Thus, the system engineering process is critical to management of availability requirements for three important reasons. First, it provides a robust process to ensure requirements are defined early in the development process. Full definition of availability requirements will address performance, the operational and maintenance environments, and the support system architecture. Second, it provides a disciplined approach to identify competing objectives and constraints, and to make logical, balanced, optimal trade-offs. Finally, the control process provides a mechanism to manage risk elements and to ensure demonstration testing is planned to confirm compliance of the system to critical RMA requirements.

2.2.2 Risk Management and Test Planning

Test activities are performed for two purposes. Testing is used to develop information on system capabilities to target corrective action and performance growth during the design and development phase. Alternatively, testing is used to determine compliance with requirements in the form of qualification or demonstration testing, most often after the configuration of the system has been fixed. Testing is the mechanism used to provide risk managers information on performance levels with respect to technical risk items. Specifically, a demonstration test is the mechanism for formal retirement of a risk item by demonstrating the system meets specified technical requirements.

Previous discussion has highlighted the importance of risk management within the systems engineering process. Risk Management controls trade-off evaluations involving critical requirements, and quantifies the risk exposure due to any uncertainty in meeting these requirements. Furthermore, the case was established that system availability is a critical system characteristic from both a performance and a cost perspective. The

following review considers the most recent information from three public sector organizations that deal routinely with high availability, long life, complex systems. These organizations include the DoD, National Aeronautics and Space Administration (NASA), and the Federal Aviation Administration (FAA).

Within the DoD, major complex system acquisition utilizes a very structured process known as the Joint Capabilities Integration and Development System (JCIDS). The JCIDS process requires regular milestone reviews that include evaluation of technical accomplishment and program risk. A program manager must successfully meet milestone review exit criteria for a program to proceed to the next phase (CJCS, 2012).

Key Performance Parameters (KPP) are attributes of a system considered critical to the development of an effective military capability. These are the few, most important characteristics of the system. Key System Attributes (KSA) are characteristics considered essential to achieving a balanced solution/approach to a system, but not critical enough to be designated a KPP. KSAs must be measurable, testable, and quantifiable. DoD requires their critical and high cost programs to include a KPP for availability as well as KSAs for reliability and ownership cost (JROC, 2012).

Operational and support costs typically exceed 50% of the total ownership cost of a program. RMA elements specify significant drivers for these costs. Program managers routinely experience constant budget and schedule pressures that increase technical and cost risk as RMA activities which occur late in the lifecycle become targets for elimination. Specific concern exists for Integrated Diagnostics for system software, and test and evaluation efforts (Department of Defense, 2005).

Within section 9.5 of the Defense Acquisition Guidebook (https://dag.dau.mil/Pages/Default.aspx, accessed on 5 Dec 2013), the DoD emphasizes that decision makers are most effectively supported by statistically defensible test results. It further states that design of experiments should be used as the structured approach to ensure statistically meaningful analyses. However, the DoD's primary reference for test and evaluation of system RMA includes a disclaimer that the handbook provides no intent that DoD test programs will produce statistically significant results, due to limitations in budget and hardware availability. Cost, schedule and operational urgency constraints could negatively impact the test and evaluation program. The goal of modified test programs is to provide essential understanding of accepted risk, and the confidence level for decisions that recommend a specific course of action. This outcome is perceived to create the most meaningful results possible (Department of Defense, 1982).

NASA has been executing a strategy over the past fifteen years to move away from rigid oversight of commercial partners. Instead, NASA implements a tailored set of guidelines that embraces mature commercial standards and best practices. This organization invests their technical energy in specifying robust, valid, testable requirements and then executing a strong risk management program to track vendor progress and manage compliance (National Aeronautical and Space Administration, 1998).

The NASA standard advocates a methodology for balancing program risk through balanced tradeoff of programmatic resources such as safety, reliability, maintainability, quality, and performance. However, consideration of the dependency between reliability

and maintainability with availability was not discussed. The availability requirement creates a dependency between reliability and maintainability, such that one cannot be arbitrarily changed without impact to the other. Furthermore, NASA states a desire to create an improved risk management program to accommodate the use of relatively unproven technology in low cost, short duration missions. This is a bold and daring vision for an organization that deals in high risk, high reward projects. However, the literature does not provide details supporting how this objective will be achieved.

The FAA viewpoint and strategy for risk and testing is the most radical and revolutionary of the three public sector organizations. This posture is founded on the belief that current RMA models and methodologies are obsolete for the FAA systems of today and tomorrow. The FAA characterizes their equipment as highly automated, high reliability systems with extremely high hardware MTBFs, and failure rates that are dominated by fast recovery software subsystems. Furthermore, the FAA guidelines state that availability is not directly testable, and therefore cannot be defined as a system requirement. They focus their requirements around MTBF, MTTR, software recovery time, and mean time between unsuccessful fault recoveries. The FAA believes these to be well understood, testable requirements (Federal Aviation Administration, 2008).

The fundamental risk that is being managed by the FAA is that of fielding a system with lower capability than the one it replaces. Thus, the FAA test program focuses entirely on the first purpose of test discussed above; one purpose of test activities is to provide failure information about the system to enable corrective action and develop system maturity. Organizationally, the FAA focuses on reliability growth, specifically driven by the identification and illumination of software errors. A developmental system

is fielded once it has been evaluated to perform better than that which it replaces. Once fielded, growth continues with persistent emphasis on identification of faults and their rapid removal.

In reviewing the RMA risk management and test philosophies of three primary public sector organizations, it is apparent that general themes are consistent, while implementation is varied. DoD and NASA strategies appear very similar to the extent that NASA does not express any concern for the effect that time and budget constraints may have on program managers. This is relevant since the DoD requirements and risk management strategies seem well defined. In spite of this, significant issues arise when schedule, budget and operational pressures impact program plans.

A possible exception to theme consistency exists within the FAA. This organization believes that statistically valid tests are impractical due to high system reliability, and the overwhelming influence of software on the system failure rate (Federal Aviation Administration, 2008). Regardless, there seems to be opportunity for ADT within all areas, if the impact of time constraints can be managed.

2.3 Demonstration Testing

A thorough test program provides a balance between testing directed at improving the system design and that which ensures the system meets critical performance requirements. The former is used to identify and remove inherent design weaknesses and is referred to as growth testing or test-analyze-fix-test (TAFT). The function of the latter is to verify compliance to performance expectations and is labeled demonstration or validation testing. This review of demonstration testing will consider its fundamental premise as well as data collection concerns and the three types of ADT.

Demonstration testing is a formalized hypothesis test in which the null hypothesis states that the system characteristic is greater than or equal to the specified requirement. The alternative hypothesis states that the system characteristic is less than some minimum specified value. When considering availability, the specified requirement is designated A_P and the minimum acceptable value is designated A_C. Formal definition of the hypothesis test would be:

$$H_0\colon\ A = A_P \tag{4}$$

$$H_1\colon\ A = A_C < A_P \tag{5}$$

As with any statistical hypothesis testing, there are two risks of making a wrong decision. First, there is a risk in rejecting a system that meets the specified requirement level. This error occurs if H_0 is rejected when it is true. This is known as the producer's risk since, the producer provided a system that met the specified requirement but was rejected, and occurs with probability α. Demonstration tests are designed such that a system meeting the specified target level will be accepted (1-α)*100% of the time. The risk associated with the minimum acceptable availability level is generally known as the consumer risk and occurs when the system does not meet the minimum requirement but is accepted anyway. This risk is realized when the null hypothesis is not rejected when H_1 is true, and occurs with probability β. Demonstration tests are designed such that a system at the minimum acceptable level will be accepted β*100% of the time [Ebeling, 2010].

A challenge in specifying a demonstration test plan is defining the values for A_P, A_C, α, and β that facilitate a test decision in a timely manner. It is apparent that A_P is specified within the documentation of the system engineering requirements loop. The

remaining three parameters should be negotiated between the producer and consumer to ensure a test framework with acceptable risk to both parties. Within the definition of specific availability targets and associated risk levels, many test plans exist. Unique specification of test characteristics appropriate for the selected ADT test strategy will define a specific test plan. Operating characteristic (OC) curves provide a graphical representation of each specific test plan. This graph plots the probability of acceptance on the y-axis against the level of the characteristic being demonstrated on the x-axis [Department of Defense, 1982]. An example of an ADT OC curve taken from Usher and Taylor (2006) is shown in Figure 4. This specific test was designed for $\alpha = \beta = 0.10$ and test hypotheses

$$H_0\colon\ A \geq A_P = 0.95 \tag{6}$$

$$H_1\colon\ A \leq A_C = 0.80 \tag{7}$$

Under this test plan, a true system availability of 0.95 has a (1-α)*100% or 90% chance of being accepted. Likewise, a true system availability of 0.80 has a β*100% or 10% probability of acceptance. This is seen in Figure 4 by locating the availability level on the x-axis and following the line vertically until the OC curve is intersected. Following the point of intersection horizontally to the y-axis will reveal the probability of acceptance.

Demonstration testing for a binary state, repairable system involves operating the system until a failure event occurs, restoring functionality to a like-new state, and then continuing operation. This cycle is shown in Figure 5. The system operates from t_0 to t_1 at which time a failure occurs. A repair action restores operation at t_2. The system operates until failure at time t_3, and is restored again at t_4. The x_i values represent the

system time-to-failure statistics, and the y_i values represent the time-to-repair. It is straight forward that the calculations are:

$$x_1 = (t_1 - t_0)\ and\ x_2 = (t_3 - t_2)$$

$$y_1 = (t_2 - t_1)\ and\ y_2 = (t_4 - t_3)$$

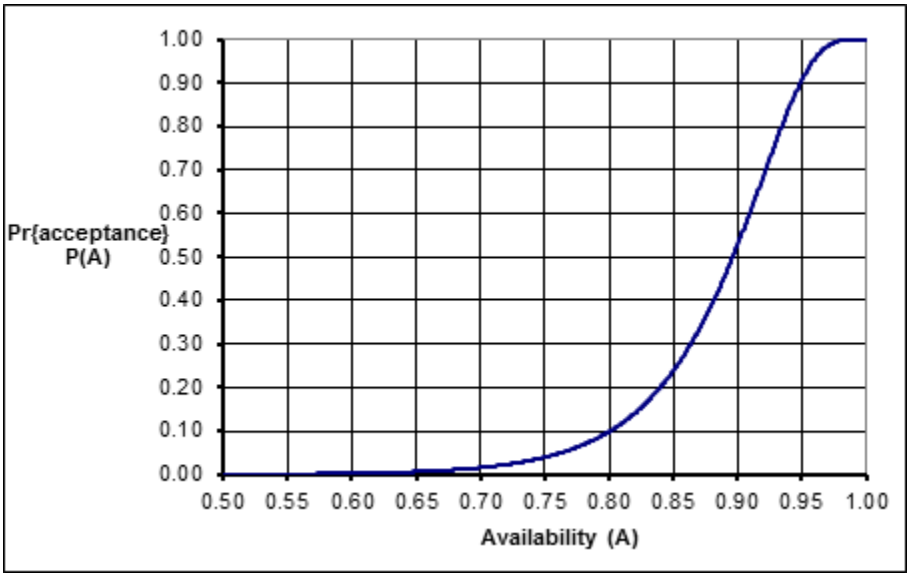

Figure 4. OC Curve for Fixed Number of Failure ADT (n*=6, z*=0.116)
[Source: Usher and Taylor, 2006]

The test metric is most often defined as the ratio of cumulative down time to cumulative operating time. Given the definition of X and Y as

$$X = \sum x_i \tag{8}$$

$$Y = \sum y_i \tag{9}$$

then, define Z as

$$Z = \frac{Y}{X} \tag{10}$$

where Z will be the test metric evaluated in all ADT approaches discussed, unless otherwise specified. In the case of sequential test plans, the test metric will be $Z(n)$ where n represents the test stage under evaluation.

Failure and repair data obtained from the process in Figure 5 are ordered pairs (x_i, y_i) collected from the same test event. Each pair represents a failure-repair cycle in the operating timeline. Alternatively, it is possible to collect this data from independent sources. In most DoD test programs, reliability demonstration and maintenance demonstration are separate test events. The test units are typically different, although they come from the same production lot. Subscripts in these data sets imply some ordering within the independent test events, but not the direct relationship between x_i and y_i stated in the former methodology. It is noted that the number of failure data produced in reliability demonstration testing does not have to equal the number of repair data produced in maintenance demonstration testing, since the tests are independent.

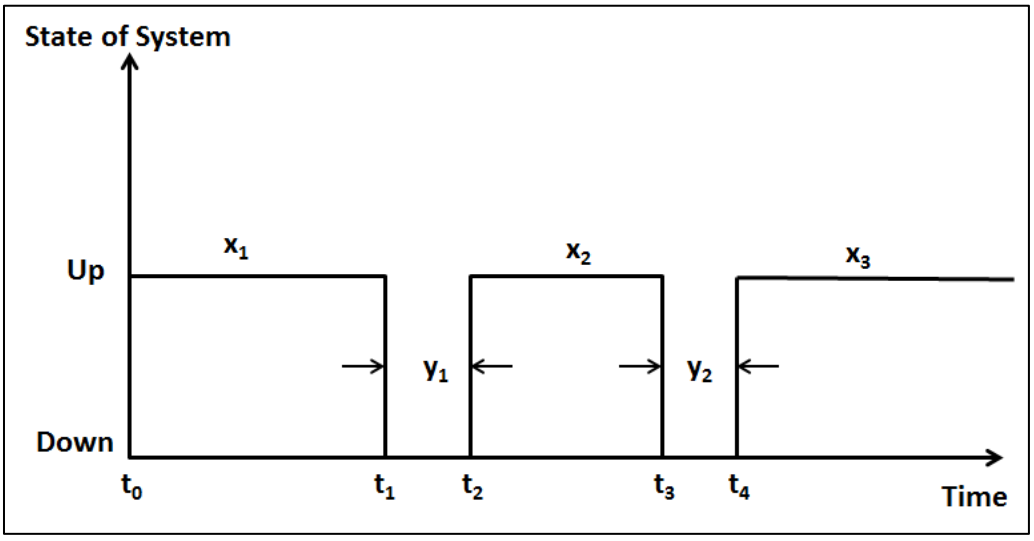

Figure 5. Demonstration Test Process

The literature uniformly assumes independence between the failure and repair distributions and their associated random variables. Thus either method should be

considered acceptable. It is interesting to note that early literature allows for failure and repair sample sizes to be different (Thompson, 1966; Gray and Schucany, 1969) while the more recent publications are based on paired data (Schafer and Takenaga, 1972; Pell, Hall and Schneider, 1978; Rise, 1979; Hällgren, 1986; Usher and Taylor, 2006). Given the independence assumption, a method has been proposed to allow a test designer to coordinate the data collected from independent reliability, maintainability, and availability demonstration testing, to minimize the cost of data collection in meeting the RMA test requirements (Fu, Yu, Zhang, and Xu, 2012). This research will focus on paired data test strategies whereby data is collected solely from a continuous ADT.

ADT plans can take one of three designs. These include fixed number of failures (FNF), fixed test time (FTT), and sequential testing (SEQ). A FNF type ADT plan specifies a number of failure-repair cycles in which data is collected. The test duration is a random variable. A FTT plan specifies test duration and evaluates the test metric at the expiration of time. In this case, the number of failure-repair cycles is a random variable. SEQ plans specify upper and lower limits which converge asymptotically as the number of failure-repair cycles *n* increase and are equal at *n* equals infinity. The test statistic is evaluated at the end of a cycle against the test limits for that cycle count. A test metric exceeding the upper limit results in a fail decision while a test metric value below the lower limit results in an accept decision. If the test metric falls between the two limits, another test cycle is run and the metric reevaluated. This continues until a decision is reached (Rise, 1979; British Institute of Standards, 1993). Both the number of failure-repair cycles and the test duration are random variables in sequential testing. Test planning can allow for truncation of a SEQ plan, although this will affect the true

producer and consumer risks levels of the test plan (Pell, Hall, and Schneider, 1978; Rise, 1979).

2.3.1 Fixed Number of Failure ADT Plans

The FNF ADT plan, also referred to as fixed sample size, is the most prevalent type of testing discussed in literature. Using this test strategy, a binary state system is placed in operation until a predetermined n number of fail/restore cycles are completed. At the end of the n^{th} restore operation, a test statistic is calculated and compared to a critical value. Based on this comparison, a pass or fail decision is made.

When the repair times and failure times are both exponentially distributed, the hypothesis test statistic is F-distributed and a function of the number of cycles in the test (Thompson, 1966; Usher and Taylor, 2006). This relationship is extended to repair times and failure times that are gamma distributed (Rise, 1979).

Considering the exponential distribution is a special case gamma distribution, research on the FNF ADT strategy has focused entirely on gamma distributed failure and repair times with one exception. A method was defined for the case of exponentially distributed failure times and lognormal distributed repair times (Gray and Schucany, 1969). However, the test statistic is not F-distributed, and definition of the elements of a fixed number of failure test plan under these conditions requires numerical evaluation. A summary of the literature for FNF ADT is provided in Table 3 with reference to parametric assumptions. Further discussion on FNF ADT is limited to gamma distributed failure times and repair times, unless otherwise noted.

The producer and consumer risk statements are fundamental to all ADT strategies. These statements are defined as:

- Producer Risk Statement: A system with a true availability of A_P will pass the test with probability 1-α%.

- Consumer Risk Statement: A system with a true availability of A_C will pass the test with probability β%.

Table 3

Summary ADT Literature and Parametric Assumptions

Author, Date	Failure Distribution	Repair Distribution
Fixed Number of Failure Strategy		
Thompson, 1966	EXP	EXP
Gray and Schucany, 1969	EXP	LogNormal
Pell, Hall and Scheider, 1978	GAMMA	GAMMA
Rise, 1979	GAMMA	GAMMA
BS 5760-10.3:1993	EXP	GAMMA
Usher and Taylor, 2006	EXP	EXP
Fixed Test Time Strategy		
Rise, 1979	GAMMA	GAMMA
Hällgren, 1986	GAMMA	GAMMA
BS 5760-10.3:1993	EXP	GAMMA
Sequential Test Strategy		
Schafer and Takenaga, 1972	EXP	EXP
Pell, Hall and Schneider, 1978	GAMMA*	GAMMA*
Rise, 1979	GAMMA	GAMMA
BS 5760-10.3:1993	EXP	GAMMA
* - Assumes shape parameter Integer and Known		

When the failure distribution and repair distribution are both Gamma distributed, the test statistic is defined as ρz where:

$$\rho = \frac{MTBF}{MTTR} = \frac{A}{(1-A)} \quad \text{and} \quad z = \frac{\sum y_i}{\sum x_i}.$$

The null hypothesis is accepted if ρz is less than z^* and rejected otherwise.

Under the hypothesis test defined by (4) and (5), the producer and consumer risk statements are written as:

$$\Pr \{ \rho Z \leq z^* \mid A = A_P, n) \geq 1 - \alpha \qquad (11)$$

$$\Pr \{ \rho Z \leq z^* \mid A = A_C, n) \leq \beta \quad (12)$$

The solution (n, z^*) that solves these two equations simultaneously is the optimum fixed number of failure test plan for A_P, A_C, α, and β. A closed form equation is also defined for the operating characteristic curve (Rise, 1979).

2.3.2 Fixed Test Time ADT Plans

The FTT ADT strategy places a binary system in operation and monitors the failure and repair times until the predetermined test time T^* elapses. At this point, the test statistic is evaluated and compared to a critical value. Under this test strategy, the number of failure and repair cycles is a random variable while the test duration is a fixed constant.

The test metric for FTT is the inherent availability *A(T)*, evaluated at the end of fixed time *T*. It is defined as:

$$A(T) = \frac{\sum_1^n x_i}{\sum_1^n x_i + \sum_1^n y_i} \quad (13)$$

where x_i are the failure times and y_i are the repair times. If *A(T)* is greater than or equal to a critical value A_{Crit}, the null hypothesis H_0 is accepted. If *A(T)* is less than A_{Crit}, then H_0 is rejected (Rise, 1979). The producer and consumer risk statements are respectively:

$$\Pr\{ A(T) < A_{Crit} \mid A_P , T) \leq \alpha \quad (14)$$

$$\Pr\{ A(T) \geq A_{Crit} \mid A_C , T) \leq \beta \quad (15)$$

Rise (1979) cites previous work that proves *A(T)* is asymptotically normally distributed. Based on this review, equations (14) and (15) are solved simultaneously to define closed form equations for test time *T*, and the critical value A_{crit}. These two values specify the FTT ADT plan. An evaluation of the accuracy of this methodology based on the normal

approximation is cited for the exponential case, but is silent on gamma distributed failure/repair times.

Several concerns were addressed for FTT strategies. One issue addresses the possibility of running a FTT ADT without a failure (Usher and Taylor, 2006). The specification of *T* in the normal approximation is expected to produce a time interval in excess of fifteen times the mean up time. However, there is a non-zero probability that the test duration can expire with no failures. In this case, it is impossible to evaluate the test statistic.

A second concern involves high reliability systems. At fifteen times the mean up time, test durations can become exceedingly long and impractical. This is exactly the concern raised by the FAA in earlier discussion. State-of-the-art systems are designed to have high reliability. In these cases, the mean up time will be large. A method was discussed for system availability greater than 0.95, but this method involves the use of nomographs which are defined for but a few parameters (Hällgren, 1986). For tests utilizing system parameters outside of this group, a set of complex equations must be numerically evaluated for specific solutions.

2.3.3 Sequential ADT Plans

SEQ testing is the final ADT planning strategy considered in this dissertation. This test strategy allows for the possibility of an early decision on whether to reject or accept the null hypothesis. This decision is based on early, significantly favorable statistical test evidence biased towards the system meeting specified performance or not. SEQ is most often seen in reliability and maintainability demonstration testing in the form of binomial testing [Ebeling, 2010).

In SEQ ADT, a rejection limit forms an upper bound, and an acceptance limit forms a lower bound. The boundary in between requires the testing to continue. As the number of failure-repair cycles increases, the limits converge and reduce the continuation region. Testing continues until one of the limits is exceeded, resulting in a decision. As such, both the number of cycles, and the total test time are random variables in sequential ADT. An example of a sequential ADT plan is shown in Figure 6 (British Standards Institution, 1993).

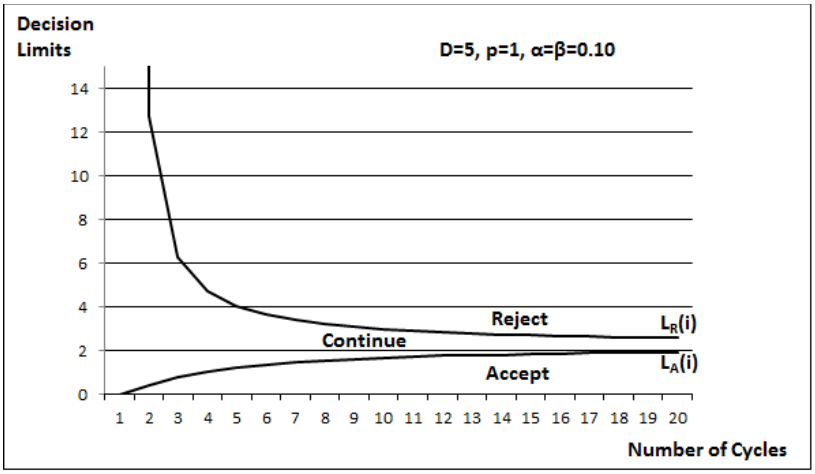

Figure 6. Sequential Availability Demonstration Test Plan

The SEQ ADT is a generalized sequential probability ratio test since the observed test metric is dependent on the previous observed value. The test metric is ρz_n , where

$$z_n = \frac{\sum_1^n y_i}{\sum_1^n x_i} \qquad n \geq 1 \tag{16}$$

and the definition of ρ is consistent with fixed number of failure testing. The value *n*, is the test cycle number. SEQ ADT will generate a set of random variables $\{Z_1, Z_2, Z_3, \ldots, Z_n\}$ across the *n* test cycles (Schafer and Takenaga, 1972).

The underlying premise of sequential testing begins with the possible outcomes of a hypothesis test. Table 4 summarizes these outcomes with respect to the validity of the null hypothesis and the test decision. If the null hypothesis is true and an accept decision is made, a correct decision has been made. A reject decision results in a Type I error which occurs with probability α. On the other hand, if the null hypothesis is false, and a reject decision is made, the correct decision is made. An accept decision results in a Type II error with probability β.

Table 4

Probability of Outcomes in a Standard Hypothesis Test

	Accept H_0	**Reject H_0**
H_0 true	Correct decision p = 1-α	Type I error p = α
H_0 false	Type II error p = β	Correct decision p = 1-β

The upper and lower limits of the sequential test strategy is based on the ratio of probabilities for an accept decision under the null and alternative hypothesis, and the ratio of probabilities for a reject decision under the null and alternative hypothesis. Consider these ratios mathematically,

$$\frac{\Pr\{Accept\ H_0\ |H_0\ true\}}{\Pr\{Accept\ H_0\ |H_1\ true\}} = \frac{1-\alpha}{\beta} \tag{17}$$

$$\frac{\Pr\{Reject\ H_0\ |H_0\ true\}}{\Pr\{Reject\ H_0\ |H_1\ true\}} = \frac{\alpha}{1-\beta} \tag{18}$$

In equation (17), if the null hypothesis is true, then the numerator should be large, and the denominator small. Thus, the value should be large. Conversely, if H_1 is true, the

denominator in equation (18) should be larger than the numerator, and this ratio should be small.

SEQ evaluates the ratio of $\rho_0 z_r$ under the null hypothesis to $\rho_1 z_r$ under the alternative hypothesis. If this value exceeds the threshold defined by equation (17), the null is accepted. If the ratio is less than the lower threshold defined by equation (18), then the null is rejected. As previously stated, testing continues until a decision is made.

Since the number of cycles in SEQ ADT is a random variable, it can be described by a distribution. The expected number of cycles is the mean of this distribution, and is referred to as the average sample number (ASN). ASN can be estimated for sequential testing. When compared to the fixed number of failure strategy, sequential testing generally has superior performance with respect to ASN. This implies that the average test duration will be shorter. However, individual tests can be considerably longer for sequential tests. Additionally, an interval exists between A_P and A_C where the ASN for sequential testing exceeds the n^* for fixed number of failures. The center of this interval is the point where the probability of accepting H_0 is equal to the probability of rejecting H_0 (Schafer and Takenaga, 1972).

Strategies have been suggested for truncating a SEQ test to prevent excessively long test durations. One method recommends limiting the cycle count to a multiple of the optimum n for fixed number of failures. The multiple is a number between 1.0 and 2.0 (Pell, Hall, and Schneider, 1978). A second method suggests a fixed time truncation, but is limited to exponentially distributed failure and repair distributions (Rise, 1979). In both cases, the true values of α and β are altered by truncation and should be considered before selecting a truncation approach.

2.4 Availability Risk Cost Modeling

Program managers are tasked with managing competing objectives while balancing resources, budget, and schedule. Prior discussion highlighted the pressure placed on DoD programs in mid and late program stages by budgetary concerns and the necessity to field systems to end users needing the capability. Decision makers are unable to make hard decisions inside this environment without robust and logical tools and support systems. Life cycle cost models are useful in situations where budget and schedule are of concern. It is important to understand how short term decisions regarding budget and schedule will impact the cost of ownership in the long run.

An excellent tool for managing such a relationship is a life cycle cost model. A proposed model for life cycle cost is (Kleyner, Sandborn, and Boyle, 2004; Kleyner and Sandborn, 2008):

$$LCC = C_{Design} + C_{Manufacturing} + C_{Validation} + C_{Warranty} + C_{Overhead} + Profit \quad (19)$$

For a long life system, this equation would have to include the sustainment and upgrade cost over the operational life of the equipment. Redefining the terms as follows:

$LCC' = LCC - Seller's\ Profit$

$C'_{Design} = C_{Design} + Design\ component\ of\ Overhead$

$C'_{Manufacturing} = C_{Manufacturing} + Manufacturing\ component\ of\ Overhead$

$C'_{Validation} = C_{Validation} + Validation\ component\ of\ Overhead$

$C_{SU} = Cost\ of\ Sustainment\ and\ Upgrade$

the life cycle cost equation becomes:

$$LCC' = C'_{Design} + C'_{Manufacturing} + C'_{Validation} + C_{Warranty} + C_{SU} \quad (20)$$

This research is specifically focused on the portion of life cycle cost related to the availability of the system. Thus, the relevant life cycle cost component for availability

would be $C'_{Validation} + C_{Warranty} + C_{SU}$. The relationship between cost and availability is shown in Figure 7 (Ebeling, 2010).

Cost of availability has two components, acquisition and support costs, and the cost of downtime. $C'_{Validation}$ is the availability component of Acquisition and Support Cost, and ($C_{Warranty} + C_{SU}$) is the component of Cost of Downtime. As availability increases, the Acquisition and Support cost element increases. This is primarily driven by increased system reliability and a more significant support infrastructure to manage mean time to repair or mean down time. This cost is also impacted by the verification test cost to demonstrate high availability. Conversely, as availability increases, the cost of downtime diminishes, whether due to fewer failure events or shorter downtimes. The minimum cost of availability occurs where these two graphs intersect.

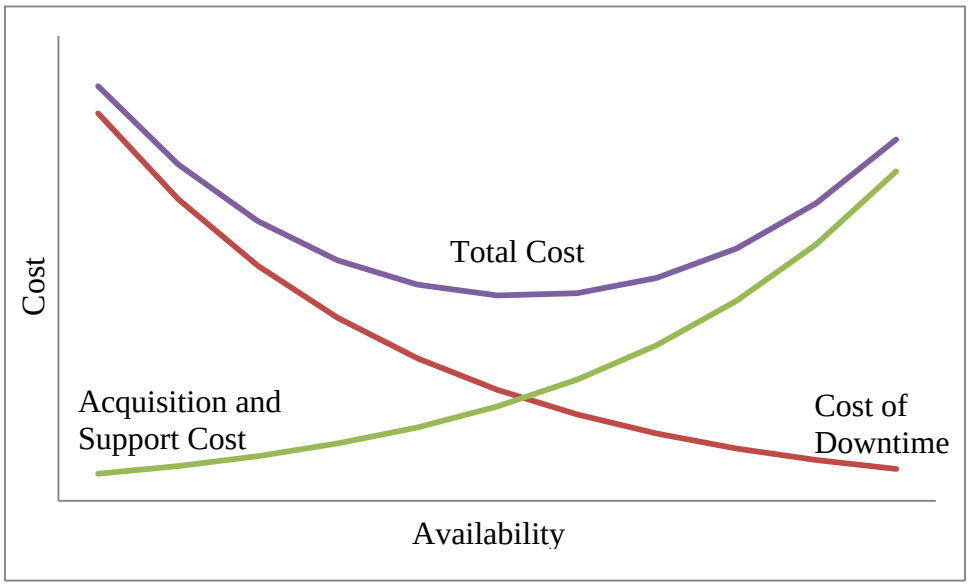

Figure 7. Cost versus Availability Curve

Acquisition and support costs for availability include the design costs to ensure availability requirements are adequately specified and designed into the system, similar costs for the specification and definition of the support infrastructure, and all test and

evaluation costs to ensure the end product is compliant to these designs. The first two elements of acquisition and support costs are managed within the requirements loop and design loop of the system engineering process. The test and evaluation is managed within the control loop, and is of concern to this research.

Cost of downtime is a part of the maintenance and sustainment of the system once it is placed in service. If a system does not comply with the defined availability requirements, then the cost of downtime will be higher than estimated. Either the number of failure events will be higher, or the maintenance associated with a failure event will be higher, or both. These outcomes are extrapolated from the causes for a non-compliant system with respect to availability; either the mean up time is lower than expected, or the mean down time is higher than expected.

Based on this relationship, a cost model is needed that captures the cost of test and evaluation, and the cost of downtime. This discussion will focus on the elements of such a cost model.

Consideration of a cost model begins with the partitioning of cost elements, and then definition of the content of each. Possible partitions include direct cost verses indirect, fixed costs verses recurring, and scalable costs that are a function of either sample size or time.

Indirect costs generally include management overhead, facility and infrastructure expenses not directly assignable to a test event, and development of capability that will be used over many products or customers (Leung and White, 1991). Direct costs include labor, equipment used directly in testing as well as its maintenance, test units, facilities and utilities, development of test plans, procedures, and reports, travel, training, and

communication with customers, and consumable items required for the test and evaluation (Leung and White, 1991; Kleyner and Sandborn, 2008; Sgarbossa and Pham, 2010).

Discussion of indirect costs suggests that these expenses are normally assigned via some activity based costing methodology as a component in the direct labor, equipment, or facility charges. As a result, there is no apparent advantage to partitioning these expenses. By capturing direct expenses, the indirect expenses are inclusive.

Fixed costs are a onetime expenditure that is the same regardless of the content of the test or the number of times a test is executed. For instance, the development of a test plan occurs once, whether the test is run one time or ten times. On the other hand, recurring costs are those that are experienced each and every time an event occurs (Leung and White, 1991; Kleyner and Sandborn, 2008).

A final consideration is the scalability of a test plan. The cost of the test can be dependent on some aspect of the plan. Leung and White (1991) discuss the dependency of test cost in terms of the number of requirements being tested. Alternatively, Sgarbossa and Pham (2010) reference test cost as a function of time. In each of these cases, the authors use these relationships to optimize a test parameter or to define termination conditions for the test. In this research, cost is a tool to help the decision maker select a desirable alternative. Hence, this scalability will be intuitively included within the recurring cost elements as the model is defined.

A general cost model has been formulated based on the content of the previously referenced literature, as synopsized in Table 5. Major categories of planning, conducting, supporting, special test equipment, and consumable items were selected from DoD

guidance on life cycle cost models (Army, 1975). This model was evaluated against project management cost estimating tools used at the Naval Surface Warfare Center, Crane Division, Crane, IN. Cost estimating tools represented component, system, and platform level cost estimates. In all cases, current test and evaluation cost estimating efforts could be mapped to this general model.

Table 5

General Test and Evaluation Cost Model Elements

		Labor		Non-Labor					
		Internal	External	Travel	Material	Facilities	OH	Non-Recurring	Recurring
Planning									
	Requirements Definition and Traceability	X	X					X	
	Test Plan Development	X	X	X				X	
	Development of Specifications, Manuals, and Procedures	X	X					X	
	Iterim Reports	X	X						X
	Final Reports	X	X						X
	Customer Interface	X	X	X					X
	Technical Community Interface	X	X	X					X
	Test Readiness Reviews	X	X	X					X
	Identification of Assets	X	X						X
	Identification of Facilities	X	X	X					X
	Training	X	X	X					X
	Schedule and/or Procurement of Assets and Facilities	X	X		X				X
Conducting									
	Transport/Set-up/Installation/Tear Down/Return Equipment	X	X	X	X	X			X
	Test Execution	X	X	X		X			X
	Simulation Execution	X	X	X		X			X
	Data Fusion and Integration	X	X	X		X			X
	Post Test Analysis/Reporting	X	X	X					X
Support									
	Engineering Support	X	X	X	X				X
	Material Analysis Support	X	X	X	X				X
	Acquisition Support	X	X		X				X
	Test Equipment/Fixture Design	X	X	X				X	
	Software Development	X	X	X				X	
	Post Failure Equipment Repair	X	X	X		X			X
Special Test Equipment									
	Test Equpment Fabrication or Procurement	X	X	X	X	X		X	
Consumable Items		X	X		X				X

The cost for downtime was more direct, given almost universal agreement within the literature. Cost elements for this category included cost for maintenance labor and cost for parts and materials associated with a failure event (Kleyner, Sandborn and Boyle, 2004; Kleyner and Sandborn, 2008; Sgarbossa and Pham, 2010). These elements coupled with the number of failure events provide the cost for downtime.

2.5 Challenges and Opportunities

Throughout the discussion in this literature review, several opportunities have been presented. These can be divided into opportunities for application of ADT, and opportunities to provide decision makers with improved knowledge of risks associated with ADT. With respect to application of ADT, the DoD acknowledges increased technical and cost risk due to budget and schedule pressure at the specific point where compliance and demonstration of capability is planned. The development of time and cost efficient ADT would provide relief to budget and schedule, and could allow testing to occur that might otherwise be cancelled with an increase in assumed risk.

Opportunity to enhance a decision maker's understanding of producer and consumer risk under time-constrained testing is an improvement in ADT methodology. Methods exist to optimize test plans based on agreement between the producer and consumer on acceptable risk levels. However, when budget and schedule pressures result in a time constraint on ADT, these risks levels are affected, usually outside the comprehension of the decision maker.

No universal guidance was found as to when it is more appropriate to use one methodology over another. Fewer samples may be needed under sequential testing than with a fixed number of failures (Ebeling, 2010). However, if the true system availability for the item under test is near the point where the chance for accepting the null hypothesis is equal to the chance for rejection, then the ASN exceeds fixed sample testing. Truncation methods were presented to counter the risk of a significantly extended test time; these methods lacked analytical assessment for magnitude of improvement or the impact to defined risk level (Pell, Hall, and Schneider, 1978; Rise 1979). Table 6

summarizes the advantages and disadvantages of the different types of test plans (UK Ministry of Defense, Part C, Chapter 40, 2011). While general statements are provided, this summary does not give insight into the quality of decisions made when comparing strategies, nor the degree of separation in timeliness of obtaining a decision.

Table 6

Advantages and Disadvantages of Demonstration Test Plans

Test Type	Advantages	Disadvantages
Fixed Number Failures/ Fixed Time	• Easier to plan and manage test resources • Maximum test time is shorter than for sequential testing	• Average number of cycles is larger than for sequential test • Very good equipment, or very bad equipment will have to complete agreed upon test duration
Sequential	• Average number of failures to reach a decision is less than fixed test plan	• Test has no maximum number of failures or time • Test resources more variable, thus schedule and budget are more difficult to manage • Continuation of test decisions require quick decision on validity of failure from one cycle to next
Truncated Sequential	• Average number of failures to reach a decision is a minimum • Test has fixed limit of time or number of failures	• Test resources more variable, thus schedule and budget are more difficult to manage • Maximum test duration exceeds fixed test plan • Only applies to EXP failure and repair distributions

A review of the literature has shown knowledge gaps that impede the effective use of ADT in a practical environment, especially for public sector decision makers.

There is a lack of documentation regarding the relative strengths and weaknesses of the existing ADT strategies. Therefore, no guidance is provided as to which strategy is most effective in a given circumstance.

The existing literature fails to address the impact of a time constraint on the test schedule. In the absence of this understanding, no support is available to a decision maker subject to a time constraint while attempting to mitigate the risk of fielding a system with sub-standard availability. This risk has extreme impact on the utility of the system in meeting mission requirements, and a potential adverse impact on life cycle cost. The need exists for a decision support tool to enable a decision maker to understand the changes to α and β as test duration changes, and make decisions based on the true risk exposure.

CHAPTER 3

COMPARATIVE EVALUATION OF AVAILABILITY DEMONSTRATION TEST STRATEGIES

3.1 Introduction

As discussed in Chapters 1 and 2, the development of a complex, large scale system relies on the systems engineering process to identify, specify, and balance critical and important characteristics during design development. The requirements loop within the systems engineering process translates user needs into specifications. The design loop translates specifications into a physical architecture suitable for production and fielding. The systems engineering control loop takes account of risk management, including test and evaluation to understand design weaknesses as well as validate system performance where required.

The Department of Defense utilizes the Joint Capabilities Integration and Development System (JCIDS) to manage the design and development of major acquisition programs. Regular milestone reviews within this process evaluate key performance parameters and key system attributes for growth and compliance before approving the program to advance into the next phase. Operational availability, mission reliability, and life cycle cost are mandated as critical system requirements which must be managed and reviewed under this process (JROC, 2012).

Availability, a function of reliability, maintainability, and logistics supportability, is a key performance driver for mission readiness and life cycle cost. Therefore, it is important for a program manager to validate the availability requirement of a developmental system as early in the lifecycle as possible. Availability demonstration testing is the tool used to validate this performance.

Availability demonstration tests can be designed using one of three possible strategies. These include fixed number of failures (FNF), fixed test time (FTT), and sequential (SEQ) testing. In each of these test strategies, a fully functional system is placed under test and operated until a failure event occurs. The system is restored to full operational capability and placed back in service. This sequence of events defines a failure-repair test cycle, and produces an ordered pair of data (x_i , y_i) where x_i is the time to failure of cycle i, and y_i is the time to repair for cycle i. Depending on the test strategy, an appropriate number of cycles are completed and a test metric is evaluated against a critical value. Based on this evaluation, the decision to accept or reject is made. It is noted that all strategies assume an equal number of failure data as repair data. Thus, all testing must terminate at the end of a full failure-repair cycle.

Under the FNF strategy, a test plan is defined by the pair (n^*, z^*) where n^* is the optimal number of failure-repair cycles and z^* is the critical value to which the test metric is compared. The test metric $z = \sum_1^n y_i / \sum_1^n x_i$ is the ratio of downtime to uptime. Given:

- the producer risk α
- the consumer risk β
- the availability under the producer risk A_P, where $U_P = (1- A_P)$
- the availability under the consumer risk A_C, where $U_C = (1- A_C)$
- Discrimination ratio $D = U_C/U_P$

BS-5760-10.3:1993 section 9.3 states the number of optimal failure-repair cycles is the minimum value $n^*=n$ that satisfies:

$$F_{1-\alpha}(2pn, 2n)\ x\ F_{1-\beta}(2n, 2pn)\ \leq D(1-U_P)(1-DU_P) \quad (21).$$

The value p is the shape parameter for the gamma failure distribution. Since this study assumes the failure distribution to be exponential, then $p=1$, and equation (21) simplifies to

$$F_{1-\alpha}(2n, 2n)\ x\ F_{1-\beta}(2n, 2n)\ \leq D(1-U_P)(1-DU_P) \quad (22).$$

The critical test statistic is defined as

$$z^*=\ F_{1-\alpha}(2n, 2n)\ x\ U_P/(1-U_P) \quad (23).$$

When using the FNF strategy, the number of test cycles is known, and the test duration is a random variable (Rise, 1979).

The FTT strategy defines a test plan represented by the pair (T^*, $A(T)^*$) where T^* is the predetermined time in which testing is terminated, and $A(T)^*$ is the critical test value. The test metric $A(T)=\sum_1^n x_i/(\sum_1^n x_i+\sum_1^n y_i)$ is the availability of the system during the test period. Continuing with the definitions of α, β, A_P, A_C, U_P, U_C, and D as stated for the FNF strategy, define:

- m_u as the mean up time or MTBF under the null hypothesis $A = A_P$
- $\lambda_{1-\alpha}$ as the 1-α upper fractile of the standard normal distribution
- $\lambda_{1-\beta}$ as the 1-β upper fractile of the standard normal distribution

When the failure distribution is exponential, then BS-5760-10.3:1993 section 9.3 defines the optimum value for T^* as

$$T^*=\ 2m_u\left\{\left(\lambda_{1-\alpha}\sqrt{(1-U_P)}+\left[\lambda_{1-\beta}(1-DU_P)\sqrt{D}/\sqrt{(1-U_P)}\right]\right)/(D-1)\right\}^2 \quad (24)$$

and the critical value $A(T)^*$ as

$$A(T)^* = 1 - \left[U_P * \frac{\left(\lambda_{1-\alpha} D(1-U_P) + \lambda_{1-\beta}\sqrt{D}(1-DU_P)\right)}{\left(\lambda_{1-\alpha}(1-U_P) + \lambda_{1-\beta}\sqrt{D}(1-DU_P)\right)}\right] \quad (25)$$

Under a FTT strategy, the number of test cycles is a random variable while the test duration is fixed (Rise, 1979).

The SEQ strategy allows for an early accept or reject decision to be made in the presence of overwhelming test evidence. Based on test parameters, an upper bound and a lower bound are defined, where both bounds are a function of the test cycle number. These bounds converge as *n* increases, and are theoretically equal at some *n* less than infinity (Schafer and Takenaga, 1972). At the end of each test cycle, the test metric $z = \sum_1^n y_i / \sum_1^n x_i$ is compared to the upper bound and lower bound. If the metric exceeds the upper bound, the system is rejected. If the metric is less than the lower bound, the system is accepted. If the metric is evaluated between these bounds, the testing continues with an additional cycle. At the completion of the next cycle, the test metric is evaluated against the new bounds. A decision is reached when the test metric exceeds one of the two bounds.

The calculations behind the SEQ strategy as defined in BS 5760-10.3:1993. Section 9.3 of the standard defines:

$$L_A(n) = \frac{D - \left[D^{1/(1+p)}\left(\frac{1-\alpha}{\beta}\right)^{1/(i+ip)}\right]}{p\left[D^{1/(1+p)}\left(\frac{1-\alpha}{\beta}\right)^{1/(i+ip)} - 1\right]} \quad \text{if} \quad n > \frac{ln\left(\frac{1-\alpha}{\beta}\right)}{p[ln(D)]} \text{ else } \quad L_A(n) = 0 \quad (26)$$

and

$$L_R(n) = \frac{D - \left[D^{1/(1+p)}\left(\frac{\alpha}{1-\beta}\right)^{1/(i+ip)}\right]}{p\left[D^{1/(1+p)}\left(\frac{\alpha}{1-\beta}\right)^{1/(i+ip)} - 1\right]} \quad \text{if} \quad n > \frac{ln\left(\frac{1-\beta}{\alpha}\right)}{p[ln(D)]} \text{ else } \quad L_R(n) = \infty \quad (27)$$

where $L_A(n)$ is the acceptance threshold and $L_R(n)$ is the rejection threshold for cycle n. The decision is made to reject the system when

$$\frac{A_P}{(1-A_P)}\frac{\sum y_i}{\sum x_i} > L_R(n) \quad (28)$$

while the decision is made to accept the system when

$$\frac{A_P}{(1-A_P)}\frac{\sum y_i}{\sum x_i} < L_A(n) \quad (29)$$

and testing continues to the next failure/repair cycle while

$$L_A(n) < \frac{A_P}{(1-A_P)}\frac{\sum y_i}{\sum x_i} < L_R(n) \quad (30).$$

In sequential testing, both the number of cycles and the test duration are random variables (Rise, 1979).

An example is provided to demonstrate the use of each of the three strategies. Consider the following producer and consumer risk statements:

- Producer Risk Statement: A system with a true availability of 0.95 will pass the test with probability 90%.
- Consumer Risk Statement: A system with a true availability of 0.80 will pass the test with probability 10%.

From engineering analysis, the mean up time of the system is estimated to be 1900 hours. Table 7 provides the values of the attributes required to define the test plans under each strategy. Optimal test plans are calculated for each strategy using these values.

Table 7

Parameters for ADT plans for Example Scenario

A	B	A_P	A_C	U_P	U_C	D	m_u	$\lambda_{1-\alpha}$	$\lambda_{1-\beta}$
0.10	0.10	0.95	0.80	0.05	0.20	4	1900	1.28	1.28

The FNF test plan is defined by the pair (n^*, z^*) using (22) and (23). The right hand side of (22) is the product of D and A_P, divided by A_C. Thus n^* is the value of n that makes the left hand side of (22) less than 4.75. Table 8 shows the values for the range of n from 1 to 8. For *n* equal to 6, the left hand side of (22) is less than the right hand side of 4.75. Therefore, $n^* = 6$.

Table 8

Calculations for optimal n^* value in FNF test strategy.

\n	1	2	3	4	5	6	7	8
$F_{1-\alpha}(2n,2n)$	9.00	4.11	3.05	2.59	2.32	2.15	2.02	1.93
$F_{1-\beta}(2n,2n)$	9.00	4.11	3.05	2.59	2.32	2.15	2.02	1.93
Product	81.00	16.89	9.30	6.71	5.38	4.62	4.08	3.72

Based on evaluation of (23) with $n = 6$, then $z^* = 0.113$. Therefore, for FNF, the optimum test plan is (6, 0.113). The test would be run for six failure-repair cycles and the system accepted if $z = \sum_1^n y_i / \sum_1^n x_i$ is evaluated to be less than 0.113.

The FTT strategy is defined by (24) and (25). Using the parameter data from Table 7 and (24), the required test duration given our producer and consumer risk statements is $T^* = 4746.5$ hours. Equation (25) defines the critical test value to be 0.8941. Thus, the optimum test plan for FTT is (4746.5, 0.8941). The test should be run for 4746.5 hours, at which time the test metric $A(T) = \sum_1^n x_i / (\sum_1^n x_i + \sum_1^n y_i)$ would be evaluated and the system accepted when $A(T) \geq 0.8941$.

The SEQ test plan includes the definition of an accept and reject boundary as defined by (26) and (27) respectively. Values on the boundary for each cycle are a function of the cycle number and are summarized in Table 9.

Table 9

Acceptance and rejection boundary for SEQ test plan in Example Scenario

N	1	2	3	4	5	6	7	8	9	10	11	12
$L_R(n)$	∞	18.392	6.757	4.773	3.955	3.509	3.228	3.036	2.895	2.788	2.704	2.636
$L_A(n)$	0.000	0.217	0.592	0.838	1.011	1.140	1.239	1.318	1.382	1.435	1.479	1.517

The boundary region defined by these points is shown graphically in Figure 8. At each cycle, the test metric $z = \sum_1^n y_i / \sum_1^n x_i$ is evaluated using the rules in (28)-(30). If equation (30) is true, then another failure-repair cycle is run and the test metric is reevaluated at the end of the new cycle using the boundary conditions for that *n* value.

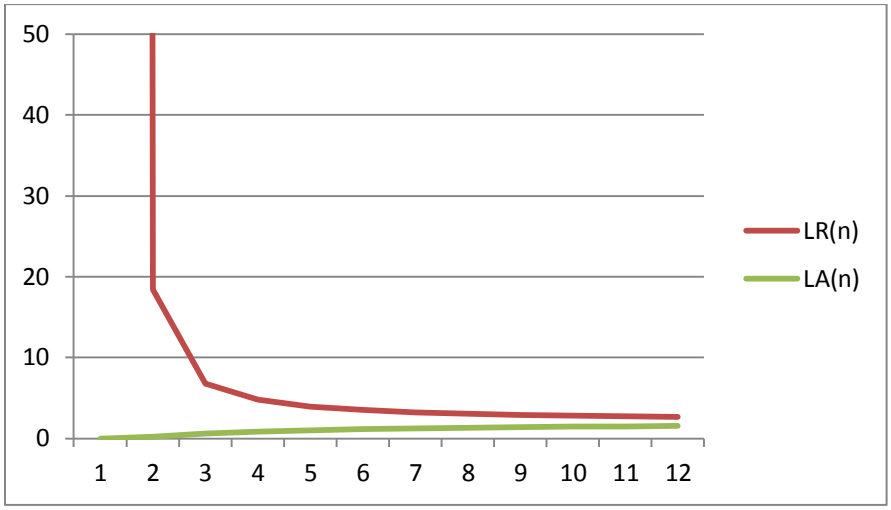

Figure 8. SEQ Test Plan for Example Scenario

The test continues until the cycle in which the test metric falls outside of the upper or lower bound. Note, the boundaries are displayed up through failure-repair cycle 12. If necessary, the boundaries can be evaluated and extended when the decision to accept or reject has not been made by the end of the 12th cycle.

Regardless of the strategy selected, validation testing requires a production representative sample, and adequate test duration to support the producer and consumer risk statements defining critical test parameters. The risk statements control the

probability of making incorrect decisions while performing the hypothesis test associated with a validation test plan. The need for production representative samples dictates the minimum point in the development cycle at which availability demonstration testing can be performed. This is typically deep in the development cycle, and most often after a production decision has been made. Therefore, the decision maker requires a test strategy that produces a quality decision in a timely manner.

A thorough review of availability demonstration testing literature reveals little guidance or insight regarding when to use a particular strategy over the other options. Evaluation of the average sample number (ASN) was compared for FNF verses SEQ. Literature suggests that a SEQ test strategy will provide a superior ASN with two caveats. In spite of the fact that the average is better, the maximum number of cycles and hence the maximum test time for a single test event is larger. Since validation test events are single test events, there is a non-trivial risk that a SEQ strategy will produce longer test durations (Rise, 1979). Additionally, given a target availability of A_P as defined by the producer risk statement, and a minimum availability A_C as defined by the consumer risk statement, there is a range between these values where the ASN for sequential testing exceeds that of the fixed number of failure strategy. This interval is centered at the point where the probability of accepting the null hypothesis is equal to the probability of rejecting the null hypothesis (Schafer and Takenaga, 1972).

Furthermore, Table 6 in chapter two summarizes the advantages and disadvantages of availability demonstration test strategies (UK Ministry of Defense, Part C, Chapter 40, 2011). Comments within the table are general regarding test durations, with disadvantages focused more on the scheduling and management of test resources.

There is no guidance on which method may be preferred, or conditions under which a preference might change.

Availability demonstration testing is an essential tool for the retirement of risk associated with fielding a system that does not meet critical availability requirements. This testing must produce quality accept or reject decisions in a timely manner. Otherwise, budgetary and schedule pressures could induce program managers or key decision makers to eliminate this vital testing in order to expedite the delivery of material to the end user. The objective of this study was to evaluate the three availability demonstration test strategies and compare their performance with respect to quality and timeliness of test recommendations.

3.2 Methodology

Availability demonstration testing on a complex, large scale system is time consuming and expensive. Evaluation of alternative testing strategies required the generation of statistically defensible results without the ability to perform direct testing on similar or like systems. Therefore, this study used a full-factorial design of experiments with test events simulated within Rockwell Automation's Arena© software.

The study assumed an exponential failure distribution with known mean, and an exponential repair distribution. The mean of the repair distribution was calculated based on the relationship

$$MTTR = MTBF * \frac{(1-A)}{A} \tag{31}$$

which is derived from algebraic manipulation of the standard availability equation. The experimental design was formulated first, and factor levels for each run were then used to define input variables and obtain test plans for the simulation sequences.

3.2.1 Design of Experiments

The experimental design was a five-factor, full-factorial design. The primary objective of the experimentation was to evaluate the significance of availability demonstration test strategy as a source of variability for measured output responses. Thus, strategy was the primary factor identified within the design. While three strategies have been discussed thus far, the design implemented four factor levels. This fourth level was a result of the assumption of equal numbers of failure and repair data.

Recall, the FTT strategy performs failure-repair test cycles until some predetermined time T^* is reached. At this point, the testing terminates. One of three possible conditions is true at the termination of testing. First, the test ends exactly at the close of the n^{th} test cycle. In this case, the n failure times, and n repair times are used in the test metric. Second, time T^* is reached during the operating period of the n^{th} cycle. In this case, (n-1) complete data pairs are available for use in evaluating A(T), and test duration is considered to be T*. In the third case, T^* is reached after the system has failed, but before operation is restored. In this event, n failure data are available, but only (n-1) repair data can be used. The test metric assumes the number of failure and repair data is the same. Thus, only the (n-1) complete data pairs can be used. The existence of multiple conditions created concern as to the exact termination conditions for a valid FTT event. This concern is further complicated when the test duration T^* lapses without completing a single failure/repair cycle. In this case, the test metric cannot be evaluated. Literature provided no insight into assessing these concerns.

Two possible solutions were considered to address this issue. The first was to terminate the testing at T*, and use (n-1) failure and (n-1) repair data for the evaluation of

A(T). This does not address the concern for test termination prior to the completion of one cycle. The second possible solution was to complete the n^{th} test cycle underway at T*, and terminate the test at the conclusion of the n^{th} restore action. Evaluation of the test metric uses the n failure and n repair data available at the close of testing. The test duration equals the total of all failure and repair times, or $\sum_1^n (x_i + y_i)$. Notice that at least one cycle must complete and therefore the concern of T* occurring during the first cycle is not an issue.

In the absence of guidance on which solution to select, the decision was made to use both alternatives as possible test strategies. In this way, the experimental design results provided insight into the most effective solution. The former solution was defined as pre-terminated Fixed Test Time and denoted as $FTT^{(-)}$ The latter solution was defined as post-terminated Fixed Test Time and denoted as $FTT^{(+)}$.

For clarity in the case of the $FTT^{(-)}$ strategy, a decision rule was established to define how test data would be used to calculate the test metric. The decision rule addressed all concerns including test termination prior to the first completed cycle. It was defined as:

- If time T^* is reached without the completion of a failure-repair cycle, the test is continued until the first cycle is complete. At that time, A(T) is evaluated with the single data pair, and the test duration is considered to be the sum of x_1 and y_1 (test duration is greater than T^*).

- If time T^* is reached during the n^{th} test cycle, A(T) is evaluated using the (n-1) complete data pairs, and the test duration is considered to be T^*(denominator of test availability is less than T^*).

To avoid the confusion of three ADT test strategies and four factor levels, the factor was identified as Method within the design of experiments and subsequent analysis of variance (ANOVA).

Additional design factors were selected from examination of the functions used to define specific test plans within the three principle strategies. Test plan optimization and evaluation was done in accordance with British Standard 5760-10.3:1993, *Reliability of systems, equipment, and components – Part 10: Guide to reliability testing, Section 10:3 Compliance test procedures for Steady-State Availability*. BS 5760-10.3:1993 methods are traceable to all literature reviewed and reported within this study. Furthermore, there are no American National Standards Institute equivalents for availability compliance testing. Thus, the British Standard provided the most logical, documented guidance for performing test simulation.

A review of section 9 within BS 5760-10.3:1993 produced a mapping of input variables and output test plan variables for each of the three strategies. This mapping is summarized in Table 10. All three test strategies require input of the producer and consumer risk levels α and β, the target availability A_P, the shape parameter of the gamma failure distribution p, and the discrimination ratio $D = (1 - A_P)/(1 - A_C)$. In addition, the FTT strategy required the mean of the failure distribution, m_u.

Table 10

Availability Demonstration Test Strategy Input and Test Plan Variables

Test Strategy	Input Variables	Test Plan Variables
Fixed Number of Failures	α, β, A_P, D, p	n^*, z_{crit}
Fixed Test Time	m_u, α, β, A_P, D, p	T^*, A(T)*
Sequential	α, β, A_P, D, p	Ac(n), Re(n)

Two assumptions were made for this study that impacted the selection of factors. First, the assumption of an exponentially distributed failure distribution has already been stated, thus p was set equal to one and dismissed for the remainder of the study. Second,

the producer and consumer risk levels were assumed to be equal. Setting these risk values at different levels implies either the producer or consumer is accepting a higher risk exposure. This must be negotiated between these partners prior to establishing a test plan. The complexity of that negotiation is outside the scope of this study. As such, the risk levels were set equal, and excursions from this posture were left for future research.

Incorporating these assumptions into the assessment of input variables, the set of parameters A_P, D, Risk Level ($\alpha = \beta$), and MTBF provided complete coverage of the parameters that effect the determination of the test plan. Hence, factors for the design of experiments were defined as Methodology, Target Availability, Discrimination Ratio, Risk Level, and MTBF. Factor levels are summarized in Table 11.

Table 11

Factor Levels for ADT Comparison Design of Experiments

	Factor Level			
Factor	Level I	Level II	Level III	Level IV
Methodology	FNF	$FTT^{(-)}$	$FTT^{(+)}$	SEQ
Target Availability	0.80	0.90	0.95	
Risk Level	0.01	0.05	0.10	
MTBF	1	100	2000	
Discrimination Ratio	2	4		

The response variables for the design of experiments were defined to measure quality and timeliness. The importance of these characteristics has already been established. The quality response is a measure of the total magnitude difference between the estimate of the probability of acceptance and the target value. This magnitude is assessed for both the producer risk statement and the consumer risk statement. The timeliness response is measured by the average simulated time to complete testing, normalized as a multiple of the MTBF. This normalization is required due to the

magnitude difference in our MTBF factor levels. More discussion on these responses will be provided after the simulation model is described.

3.2.2 Arena® Simulation Model

Arena® simulation software is a discrete event simulation tool that allows for the definition of a stochastic process, and then uses Monte Carlo simulation to assess the performance characteristics of the process. In this study, an availability demonstration test process was defined for each of the four methods being assessed. A simulated test was performed 10000 times, and the percentage of times an acceptance decision was made as well as the average time it took to make the decision were recorded.

Input variables included the five factor levels of our design of experiments, as well as the parameters of the test plan associated with the experimental run. The factor levels for a given run were defined; these levels were imported into an Excel™ spreadsheet which calculated the test plan variables for the FNF and FTT strategies. The SEQ test boundaries were generated inside the simulation model. The experimental design run factor levels and resulting test plan parameters were then read into Arena® from an Excel™ based input file. Based on the level of the Method factor, Arena® branched to the appropriate model for the test strategy and performed the Monte Carlo simulation. Output responses were then read to an Excel™ based output file for analysis.

The response variables for the design of experiments analysis are considered in greater detail. An availability demonstration test is based on a producer risk statement, and a consumer risk statement. These statements are defined as:

- Producer Risk Statement: A system with a true availability of A_P will pass the test with probability 1-α%.

- Consumer Risk Statement: A system with a true availability of A_C will pass the test with probability β%.

There is one, and only one test plan under each strategy that will produce the probabilities defined within both statements simultaneously. Thus, if a test plan is generated based on these definitions, then execution of that plan over a large sample should produce similar results. Therefore, a good measure of quality would assess how closely the actual probabilities obtained from simulation of the test plan matched the predicted probabilities. Output included the simulated probability of acceptance under the producer risk statement, the probability of acceptance under the consumer risk statement, and the average time required to complete the test.

A simulation run accepted an input of the test strategy, the target availability (producer risk), the MTBF, and the discrimination ratio. Based on the test strategy, the simulation model branched to a module that performed FNF, $FTT^{(-)}$, $FTT^{(+)}$, or SEQ testing. An index variable managed the number of cycles, and failure and repair data were generated based on the input MTBF and calculated MTTR. A single test event was concluded with the decision to accept or reject. Global variables were used to track the number of accept decisions and the total test time, which equaled the sum of all failure data and repair data. Once a test decision was made, local variables were reset and the simulation looped to begin a new test. Each simulation run produced 10000 replicated test events. At the close of the final replication, the Pr{accepting the system | $A = A_P$} was estimated by the number of accept decisions divided by 10000. The total test time was stored and all local variables were reset.

Using the discrimination ratio and the target availability, the availability under the consumer risk statement was calculated. The process was repeated with the availability

level set at the new value. At the conclusion of 10000 replications, the second simulated probability, Pr{accepting the system | A = A_C}, was calculated in the same manner explained above. Then, average test time was defined by dividing the accumulated test time by 20000. Recall that the total test time was accumulated across the all replications with A = A_P and A = A_c. Output from the simulation run included the two probability estimates and the average test time.

The quality response for the design of experiments was the total deviation of the output probabilities from their theoretical equivalents. When A = A_P, the probability of acceptance is $1 - \alpha$. When A = A_C, the probability of acceptance is β. Therefore, the quality response was defined as the sum of the ABS[(1-α) - Pr{accept system | A = A_P}] and the ABS[β - Pr{accept system | A = A_C}]. The timeliness response was the output average test time normalized by the MTBF of the failure distribution. Therefore, T = (average test time)/MTBF.

Appendix A provides sample documentation concerning the design of experiments and simulation modeling contained within this study. Additionally, this appendix contains an output summary for the details of the analysis in section 3.3 to follow.

3.3 Analysis and Results

A total of two replicates for a 4x2x3x3x3 five-factor, full-factorial design was run to assess the mean difference in availability demonstration test strategies with respect to quality and timeliness. Factors included Availability, Discrimination Ratio, MTBF, Methodology, and Risk Level. Output responses included quality and timeliness as

defined in section 3.2.2. Data was entered into MiniTab® version 16 for analysis. The initial ANOVA model included all main effects, and all possible interactions.

3.3.1 The Quality Response

The quality response was evaluated first; a review of residuals confirmed that the response data met the normality requirements (Figure C-2, Appendix C). A review of the ANOVA table quickly revealed that the MTBF main effect, and all but three interaction terms involving MTBF were not significant; reference Table C-2 in Appendix C. The three interaction terms with MTBF that were significant had the lowest F-values for all significant terms. The decision was made to remove MTBF from the model.

The revised model included main effects and all interactions for factors Availability, Discrimination Ratio, Methodology, and Risk Level. Residual analysis again confirmed the normality assumption (Figure C-3, Appendix C). Under the new model, all terms in the ANOVA table were significant. Table 12 presents the ANOVA table for the model discussed, with sources ordered in descending F-ratio value. The F-ratios experienced significant drops in magnitude after the first two terms. In fact, the sum of the F values for the remaining terms was only 60% of the second largest term. As such, analysis focused on the first two terms of the ANOVA table. The third term was added to consideration as it was the interaction of the two most significant terms. This significant interaction is evaluated first.

When considering the interaction between the Discrimination Ratio and Method main effects, the two FTT methods were sensitive to the level of the Discrimination Ratio. Under FNF and SEQ strategies, there was little difference in performance across the high and low levels for D. However, Figure 9 shows a significant increase in mean

error for the FTT methods when Discrimination Ratio was at the high level. The Tukey test for the interaction shows that all factor level combinations are significantly different. The two lowest mean response values are for Method 1 at the different D levels. The next two lowest mean response values are for Method 4. This result indicates that Method 1 is preferred over the other three, although it is sensitive to the level of D in the ADT design.

Table 12

Analysis of Variance table for Availability Demonstration Test Quality Response – (factors include A0, D, Method, RiskLevel)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
D	1	0.580734	0.580734	0.580734	55986.48	0.000
Method	3	1.710063	1.710063	0.570021	54953.67	0.000
D*Method	3	0.475943	0.475943	0.158648	15294.65	0.000
RiskLevel	2	0.198814	0.198814	0.099407	9583.48	0.000
A0	2	0.068842	0.068842	0.034421	3318.42	0.000
RiskLevel*Method	6	0.068475	0.068475	0.011413	1100.24	0.000
A0*D*RiskLevel	4	0.025076	0.025076	0.006269	604.36	0.000
A0*RiskLevel*Method	12	0.067655	0.067655	0.005638	543.53	0.000
A0*D*RiskLevel*Method	12	0.064985	0.064985	0.005415	522.08	0.000
A0*Method	6	0.020521	0.020521	0.00342	329.72	0.000
A0*D	2	0.006528	0.006528	0.003264	314.69	0.000
A0*RiskLevel	4	0.010537	0.010537	0.002634	253.95	0.000
D*RiskLevel*Method	6	0.014741	0.014741	0.002457	236.86	0.000
A0*D*Method	6	0.012657	0.012657	0.002109	203.36	0.000
D*RiskLevel	2	0.001138	0.001138	0.000569	54.85	0.000
Error	360	0.003734	0.003734	0.00001		
Total	431	3.330442				

The significant main effect terms were the Discrimination Ratio and the Method. Main effects plots for these factors are shown in Figure 10. The Tukey Test for sample means revealed that the difference between levels in the Discrimination Ratio factor was significant (Table C-4, Appendix C). Under the low level, the mean response was less

than half that of the high level. Therefore, over all test methods, as the separation between the target availability and the minimum acceptable availability gets larger, the quality of the test degrades.

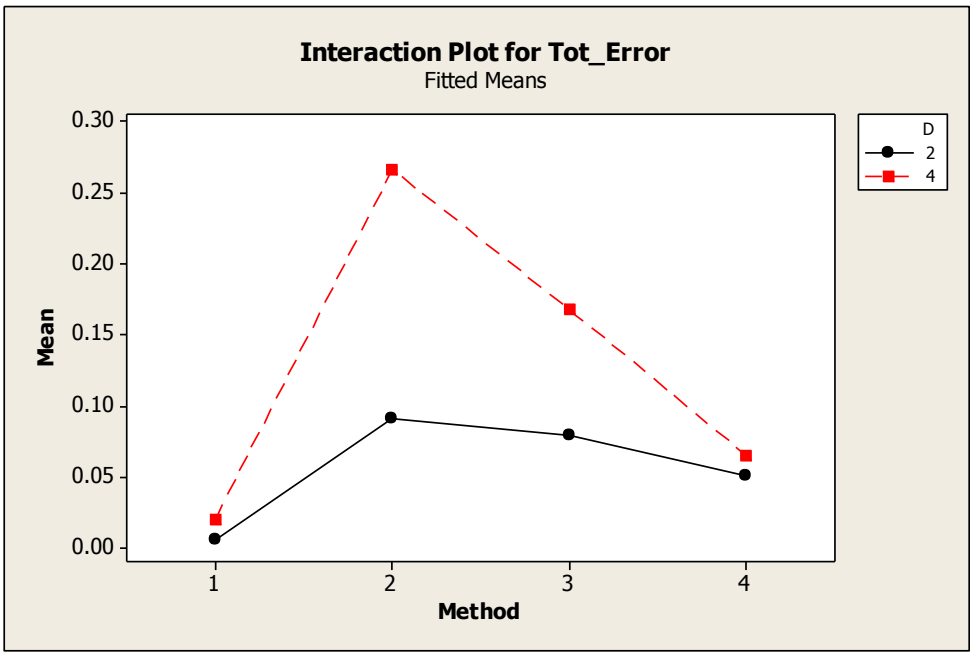

Figure 9. Interaction Plot for Discrimination Ratio and Method (Quality Response)

Table C-5 in Appendix C summarizes the Tukey Test for Method. All factor level pairs are significantly different. The best performance resulted from the FNF test strategy, followed by the SEQ strategy. The two FTT strategies performed worst, with post-terminated FTT ($FTT^{(+)}$) outperforming the pre-terminated FTT ($FTT^{(-)}$).

Given that a smaller mean is better, the following summary of results for the quality response was noted:

- When considering the interaction between Method and Discrimination Ratio, the difference between mean error for Discrimination Ratio levels was significantly larger for both FTT methods, than for either FNF or SEQ strategies. For the $FTT^{(-)}$ method, as Discrimination Ratio increased from 2 to 4, the mean error increase by nearly 300%. Reference Table C-6 in Appendix C.

- All factor level pairs for Method were significantly different. The preference for Method was FNF, SEQ, $FTT^{(+)}$ and $FTT^{(-)}$. The mean error for the SEQ strategy was 450% larger than for FNF. The mean error for the FTT methods was an order of magnitude greater than for FNF. Thus the fixed number of failure strategy performed most like its theoretical results in the simulation runs. Reference Table C-5 in Appendix C.

- A discrimination ratio of 2 had a smaller mean error than 4. This would imply that as the separation between the availability levels specified in the producer and consumer risk become larger, the simulated performance was less precise. Reference Table C-4 in Appendix C.

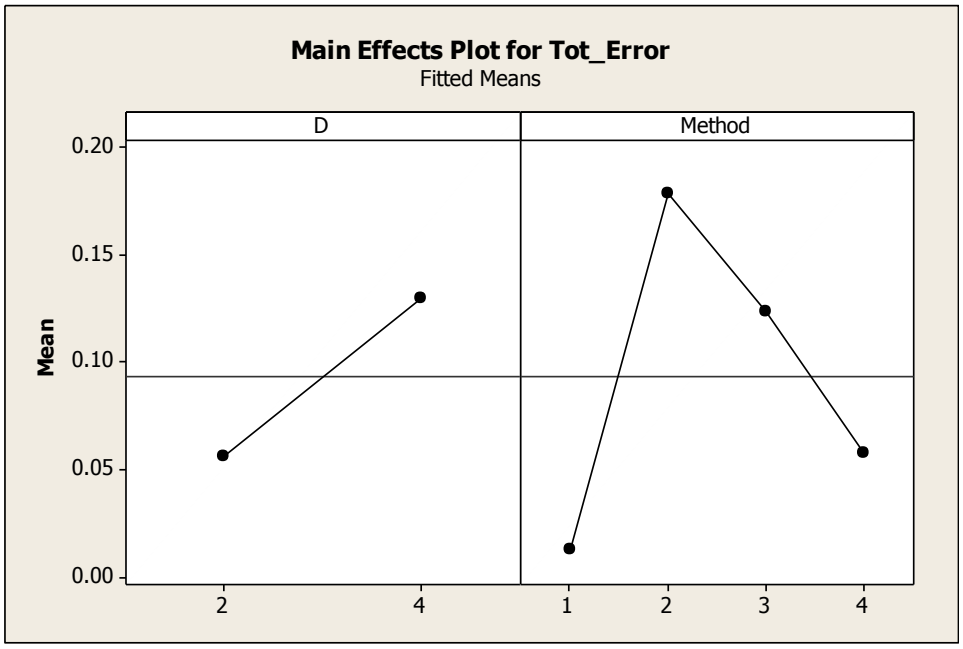

Figure 10. Main Effects Plot for Discrimination Ratio and Method (Quality Response)

3.3.2 The Timeliness Response

The base ANOVA model for the timeliness response was defined to include all main effects and all interactions for the five factors. Analysis of residuals indicated a lack of linearity in the normal probability plot, as well as non-constant variance across the fitted values. See Table C-4 in Appendix C. As a result, the normality assumption could not be substantiated. The Box-Cox transformation method was used within

MiniTab® to estimate an optimum power value of 0.17. The transformed data had significantly better linearity and constant variance as shown in Table C-5 in Appendix C. ANOVA was performed on the transformed data.

Table C-7 in Appendix C presents the ANOVA table for analysis of the transformed data. Similar to observations in the Quality analysis, the main effect MTBF and all but one of its interactions are not significant. The single interaction that is significant has the lowest F value of significant terms. Therefore, MTBF was removed from the model and the ANOVA was recalculated. These results are shown in Table 13.

Table 13

Analysis of Variance table for Availability Demonstration Test Transformed Timeliness Response – (factors include A0, D, Method, RiskLevel)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
D	1	17.079520	17.079520	17.079520	29155733.2	0.0000
RiskLevel	2	4.982010	4.982010	2.491000	4252288.2	0.0000
Method	3	3.982750	3.982750	1.327580	2266264.0	0.0000
D*Method	3	0.817140	0.817140	0.272380	464965.8	0.0000
D*RiskLevel	2	0.327510	0.327510	0.163760	279540.6	0.0000
A0*D	2	0.196680	0.196680	0.098340	167875.2	0.0000
A0*Method	6	0.348850	0.348850	0.058140	99249.8	0.0000
A0	2	0.071050	0.071050	0.035530	60646.1	0.0000
RiskLevel*Method	6	0.142990	0.142990	0.023830	40681.9	0.0000
A0*RiskLevel	4	0.069010	0.069010	0.017250	29452.3	0.0000
D*RiskLevel*Method	6	0.052720	0.052720	0.008790	15000.7	0.0000
A0*D*RiskLevel	4	0.033790	0.033790	0.008450	14418.8	0.0000
A0*D*Method	6	0.019840	0.019840	0.003310	5645.1	0.0000
A0*RiskLevel*Method	12	0.038460	0.038460	0.003200	5470.9	0.0000
A0*D*RiskLevel*Method	12	0.033200	0.033200	0.002770	4723.6	0.0000
Error	360	0.000210	0.000210	0.000000		
Total	431	28.195740				

The magnitude of the F ratio drops significantly after the third term. Therefore, analysis was focused on those three terms, the main effect terms for Discrimination

Ratio, Risk Level, and Method. The main effects plots for these factors are shown in Figure 11. As the Discrimination Ratio increases from its low level to its high level, the mean measure of timeliness decreases by over 20%. The same trend is seen as the Risk Level increases from its lowest level to its highest level. These effects were predictable, and can be explained by examining the mathematical equations for each test strategy within Section 9 of BS 5760-10.3:1993.

Within FNF, the Risk Level and Discrimination Ratio vary inversely with the number of failure-repair cycles required in the optimum test plan. Therefore, as these variables decrease in magnitude, the number of cycle n* increases. The result is an increase in the expected test duration. Similarly, as Discrimination Ratio and Risk Level decrease in magnitude, the value T* increases for the FTT strategy. As T* increases, the expected test duration will increase.

The dynamic for the SEQ strategy is more subtle. As the risk level increases, the decision boundaries collapse towards each other. As this happens, the probability of making a decision within each cycle increases. As this probability increases, the likelihood of requiring additional cycles decreases as does the expected test duration. Therefore, as Risk Level increases, the duration of the test decreases. Risk Level has an inverse relationship with the expected test time for all three strategies. While there is no obvious relationship between the Discrimination Ratio and expected test duration in the SEQ strategy, there is an obvious inverse relationship seen in the other two strategies. The existence of a relationship in two of the three strategies is sufficient to establish significance in the factor level effect.

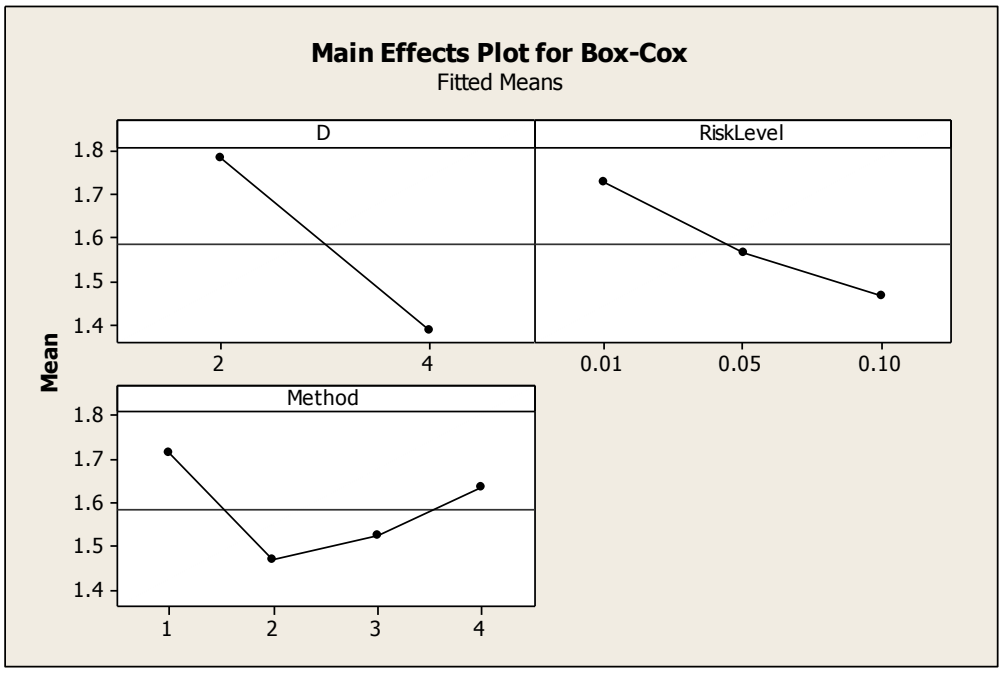

Figure 11. Main Effect Plot for Discrimination Ratio, Risk Level and Method (Transformed Timeliness Response)

Table C-10 in Appendix C summarizes the pairwise comparison for the Method factor. All factor pairs are significantly different. As the factor effect plot in Figure 11 displays, Method 2 has optimum performance, followed by Method 3, then Method 4 and finally Method 1. This translates to $FFT^{(-)}$, $FTT^{(+)}$, SEQ, and then FNF. Ironically, this is precisely the opposite order as seen under the quality response. So strategies that are strongest in quality are weakest in timeliness, and vice versa. The SEQ strategy occupies middle ground in both measures.

3.3.3 Overall Results

The results for the study are summarized in Table 14. The factor Method was statistically significant in analysis of both output responses, Quality and Timeliness. This suggests that the ADT strategy selected will have an impact on the Quality of the accept/reject decision made, and will have an impact on the time to complete the test.

Table 14

Summary of Results for Comparison of ADT Strategies

Response	Order of Preference (Best → Worst)				Significant Interactions
Quality	FNF	SEQ	$FTT^{(+)}$	$FTT^{(-)}$	Method-Discrimination Ratio
% of best	100%	450%	950%	1370%	
Timeliness	$FTT^{(-)}$	$FTT^{(+)}$	SEQ	FNF	
% of best	100%	124%	188%	250%	

The order of preference for method was exactly opposite for the evaluation on Quality as it was for the evaluation of Timeliness. The relative performance of each factor level is shown in Table 14. These percentages for timeliness are based on the untransformed data as summarized in Table C-11 in Appendix C. The best choice for the Quality analysis was the last choice under the Timeliness analysis, and vice versa. A significant interaction between Method and Discrimination Ratio was established in the ANOVA for the Quality response. This interaction suggests sensitivity in the FTT strategy to the discrimination ratio of the test. As the separation between target and minimum acceptable availability increases, the quality of decision making at these points diminishes.

3.4 Conclusions

Decision makers and program managers directing the development of large scale, complex systems require the timely validation of availability requirements subject to the defined producer and consumer risk levels specified for a test event. This study evaluated three availability demonstration test strategies with respect to Quality of accept/reject decisions and the Timeliness of test plans.

A full factorial design of experiments was defined for five factors, Method, Target Availability, Discrimination Ratio, MTBF, and Risk Level. The design included two

replicates. Response values for Quality and Timeliness were simulated within Arena® software based on no less than 10000 replications. ANOVA was performed on both the Quality and the Timeliness response data, based on the confirmation of the normality assumption via residual analysis.

Method was determined to be significant with respect to both Quality and Timeliness. However, the order of preference for the four methods was reversed between the two analyses. When considering the quality response, the FNF strategy was the best in performance. The SEQ strategy has a mean error 350% larger than FNF. The $FTT^{(+)}$ and $FTT^{(-)}$ strategies had mean errors that were 850% and 1270% larger respectively.

When considering the timeliness response, the FTT strategy was best in performance. The pre-truncated method of FTT had the strongest overall performance with respect to timeliness. The post-truncated method took 24% longer on average. The SEQ strategy took 88% longer than $FTT^{(-)}$, and the FNF strategy had the worst performance at 2.5 times longer on average.

Note, the ANOVA on the quality response data indicated an interaction between the Method factor and the Discrimination Ratio factor. This interaction was due to a significantly larger mean error when the discrimination ratio was at the high level under FTT testing, than when it was at its low level. The difference was much smaller for FNF and SEQ strategies.

A decision maker must trade off the value of the quality and the timeliness in obtaining the decision when deciding which strategy to use. If deviations in the defined producer and consumer risk levels are acceptable, and time is constrained, then the Fixed Test Time strategy should be chosen. When choosing FTT, pre-termination has better

performance than post-termination at the expense of quality with respect to the risk levels. However, the test designer must be conscious of the Discrimination Ratio. A large separation between the target availability and a minimum acceptable availability will adversely impact the quality of the decisions, as shown by the significant interaction displayed in Figure 9.

On the other hand, if the program is sensitive to deviations in the α and β risk levels, but has few or no time constraints, then the performance of the Fixed Number of Failure strategy is superior and should be chosen.

Reality seldom provides the decision maker with circumstances that support an obvious answer. The sequential test strategy provides a balance between quality and timeliness. It has better performance than FTT in the quality evaluation, and better timeliness than FNF. However, the test designer must remember that while the average sample number of the test is likely to be smaller, the maximum possible test duration is larger.

These general conclusions provide improved guidelines over those available in literature. A better solution would provide the decision maker with a mechanism to evaluate the producer and consumer risk levels as the time to complete a test is modulated. Such an approach is formulated in the next chapter.

3.5 Future Research

The objective of this study was to evaluate the significance of availability demonstration test methodologies in the variation seen in test quality and timeliness. This insight provides decision makers with guidance on which test strategy to select

under differing conditions. There are four major areas where this research can be expanded to improve this guidance.

Sequential demonstration test techniques provide potential to reduce test duration. However, while the mean of test time is lower than its alternatives, the larger variance can be problematic to a constrained budget and schedule. The maximum test time is definitely larger based on literature (Schafer and Takenaga, 1972). The literature provided no discussion on the evaluation of the probability that test duration under the sequential strategy will exceed the test duration under the alternative strategies. Understanding the average sample number is good information, but would be far more useful if paired with the estimated probability discussed above.

Further expansion on the Sequential test strategy could be accomplished by looking at truncation methods which limit the test time. These approaches retain the advantage of a shorter average test length, while limiting the disadvantage of a significantly larger maximum test time. The use of design of experiments to manage a simulation analysis aimed at evaluating the effect of truncation on producer and consumer risk levels would be valuable.

The simulation study summarized by this discussion focused on an exponential failure distribution and an exponential repair distribution. The strategies which define the Method factor levels are defined in terms of an exponential failure distribution and a gamma repair distribution. Given the exponential has limited applicability to practical real world repair processes, expanding the study to the more powerful gamma repair assumption would be useful to expanding decision guidance.

A final consideration involves the assumption of equal producer and consumer risk. The effect of deviating from this assumption is not obvious from the results of this study. This assumption restricts the trade space between the producer and consumer when negotiating the most desirable risk levels for a verification test. As such, knowledge regarding the impact of changes in risk levels to the variation in the responses quality and timeliness are relevant and significant.

CHAPTER 4

AVAILABILITY DEMONSTRATION TESTING WITH TIME CONSTRAINT

4.1 Introduction

Availability is a critical measure of system performance and dependability. It has a direct relationship with the probability of accomplishing the mission of the system, as well as the life cycle cost of the item. The consumer of large-scale complex systems must consider and mitigate the risk of acquiring and fielding a system that does not meet their availability requirements. Otherwise, the consumer is put at risk of failing to accomplish the purpose of the system or incurring excessive cost of ownership in fielding and sustaining the product.

Availability demonstration testing is a risk mitigation tool that enables a consumer to verify system performance against specified requirements prior to placing it into service. Three well defined ADT strategies have been presented, namely, fixed number of failures, fixed test time, and sequential testing. In Chapter 3, these strategies were evaluated with respect to quality of decision and timeliness to complete testing. A full-factorial design of experiments was completed with analysis of variance run on the resulting data. Conclusions from the ANOVA indicated that the FNF strategy performed better in quality of decision as measured by deviation from producer and consumer risk levels defined for a test plan. Alternatively, the FTT strategy provided superior performance in terms of timeliness as measured by average time to complete testing.

The ADT process requires a fixed product configuration in order to validate performance. As such, the timing of validation testing typically occurs after the delivery of low rate production units at the start of manufacturing (Department of Defense, 2008). This timing within the product lifecycle places significant pressure on the program manager or decision maker, as schedule and budget slippages from the design and development phases increases pressure to accelerate delivery to the end user (Department of Defense, 2005). Regardless of whether the pressure originates from a need to recover a schedule slippage or to reduce test cost to deal with budget overruns, the end result can lead the decision maker to reduce the time allocated to run the demonstration testing. In the worst case, testing may be eliminated altogether.

This study will show that a reduction in test duration while holding availability targets constant will have a direct negative effect on the producer and consumer risk levels for a given ADT plan. It is assumed that the decision maker desires to understand the magnitude of change in the risk levels in order to effectively manage the overall risk exposure to the program. As such, an approach is developed and presented which specifies the relationship between producer risk level, consumer risk level, and the critical test value when a time constraint is place on an ADT event. The cost of availability risk mitigation will be used as the context in which producer and consumer risk levels will be balanced to most effectively meet the needs of the decision maker.

4.2 Time-Constrained Testing

Strategies to define ADT plans are well documented and have been presented within the literature review in Chapter 2. Depending on the specific strategy selected, the duration of the test can be fixed or a random variable. These strategies assume that the

decision maker has the time available to execute the optimized test plan. However, the case has been made that real-world budget and schedule pressure can create time constraints on a program's test and evaluation schedule.

If time constraints impact the test plan in such a way to prevent full execution of the test plan, the true producer and consumer risks for the test will be greater than those stated within the risk statements used to define the original plan. The probability of rejecting a system that meets user requirements will increase simultaneously with an increase in the probability of accepting a system that fails to meet user requirements. This increase in risk exposure is undesirable to both the producer and consumer. In order to effectively manage the program risk, the decision maker must understand the true level of producer and consumer risk when a time constraint is imposed.

When terminating the FNF and FTT strategies earlier than specified, the decision to accept or reject can be made with the understanding that the true α and β values will be greater than specified. However, under the SEQ strategy, if a fixed termination point is defined, it is possible for the final decision of "continue to test" to prevail at the terminal point of testing. In this event, the test concludes with a no decision. This is an unacceptable outcome and would not be considered as an alternative to the decision maker. Therefore, this study does not consider the SEQ strategy in the event of a time constraint. The study will limit its consideration to FNF and FTT strategies.

4.2.1 Time Constraint on the Fixed Number of Failure Test Strategy

The producer and consumer risk statements have been specified in equations (11) and (12). The mathematical equivalent of these statements for the fixed number of failure

strategy with exponential failure and repair distributions is defined as (Usher and Taylor, 2006):

$$\int_0^{z^*} \frac{\Gamma\left(n^{*2}\right)}{[\Gamma(n^*)]^2}\left(\frac{A_P}{1-A_P}\right)\frac{x^{n^*-1}}{(1+[A_P/(1-A_P)]x)^2}dx \geq 1-\alpha \tag{32}$$

$$\int_0^{z^*} \frac{\Gamma\left(n^{*2}\right)}{[\Gamma(n^*)]^2}\left(\frac{A_C}{1-A_C}\right)\frac{x^{n^*-1}}{(1+[A_C/(1-A_C)]x)^2}dx \leq \beta \tag{33}$$

Consider the graphs in Figure 12. The top graph reflects equation (32). The area under the curve to the right of z_1 is equal to α. Therefore, the area under the curve for $z_1 < z^*$ will be less than α, making (32) true.

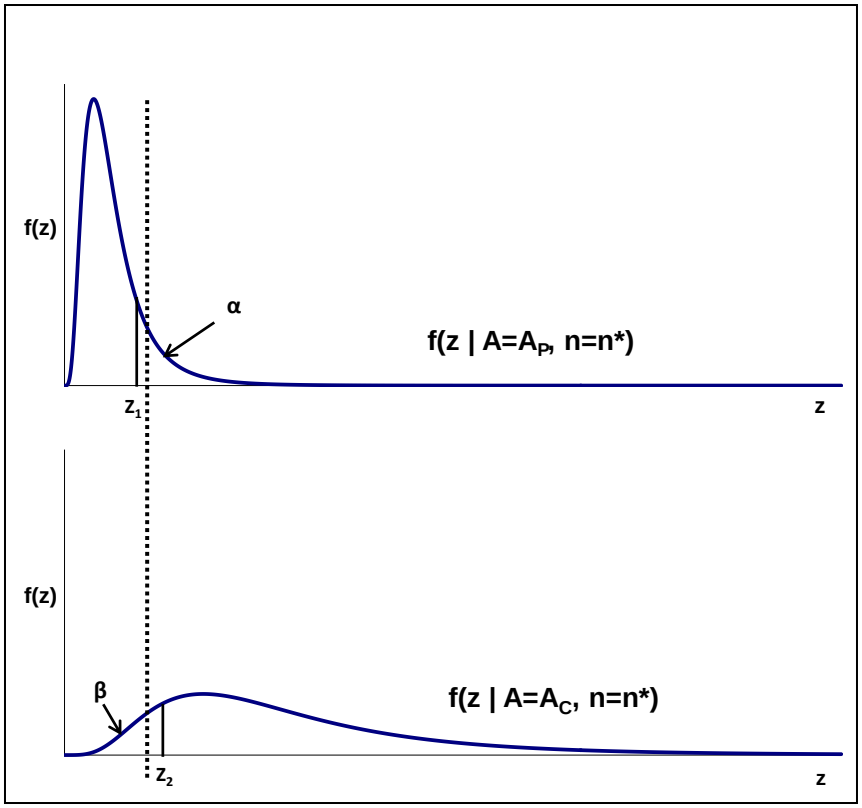

Figure 12. Graphical Representation of Optimal FNF Test Plan under Producer and Consumer Risk Statements

The graph at the bottom of Figure 12 represents equation (33). The area to the left of z_2 is equal to β. Thus, the area under the curve to the left of any $z^* < z_2$ will be less

than β. Given the dashed line crosses both graphs at the same value for z^*, anytime the dashed line is between z_1 and z_2, the conditions of (32) and (33) are met. Given A_P, A_C, α, and β are known and defined, an optimal test plan (n^*, z^*) for FNF is any simultaneous solution to these two equations. This can be either a unique test plan when $z^* = z_1 = z_2$, or several test plans when $z_1 \leq z^* \leq z_2$ and $z_1 \neq z_2$.

Each of the n^* cycles is the sum of an operating time to failure x_i, and the time to restore system operation y_i. In steady state, the expected value of x_i is the mean of the failure distribution, or the MTBF. The expected value of y_i is the mean of the repair distribution, or the MTTR. Thus, the expected length of a single cycle is MTBF+MTTR. Furthermore, the expected value for the length of the entire test is

$$E[Test\ Duration] = n^*(MTBF + MTTR) \tag{34}$$

Equation (34) defines the point estimate for the random variable representing total test time under the test plan (n^*, z^*).

Now, consider the impact of a time constraint designated by the program manager or decision maker. If the expected test time defined by equation (34) exceeds the time available for the demonstration event, then a shorter test duration *T'* may be specified by the decision maker. Given a stated value for *T'< E[Test Duration]*, then a sub-optimum number of cycles *n'* is defined as

$$T' = n'(MTBF + MTTR) \tag{35}$$

and the allowed number of cycles is defined as

$$n' = T'/(MTBF + MTTR) \tag{36}$$

Since *n'* represents the number of cycles, it must be an integer value. It will be up to the

decision maker whether the value is rounded up or down. Regardless of which direction n' is rounded, by definition, $n' < n^*$.

In section 3.1, it was stated that n^* is the minimum number that satisfies equation (21). Therefore, n' cannot satisfy equation (21), and there can be no simultaneous solution to equations (32) and (33) for $n = n'$ since it is less than n^*. Figure 13 shows this non-optimal condition. When $n' < n^*$, $z_1 > z2$ and there is no value for z which satisfies equations (32) and (33) simultaneously. Therefore, given $n' < n^*$, for any value of z, α' is defined as the producer risk under n' equal to the area under the curve to the right of z with $A = A_P$. Likewise, β' is defined as the consumer risk under n' and equals the area under the curve to the left of z with $A = A_C$.

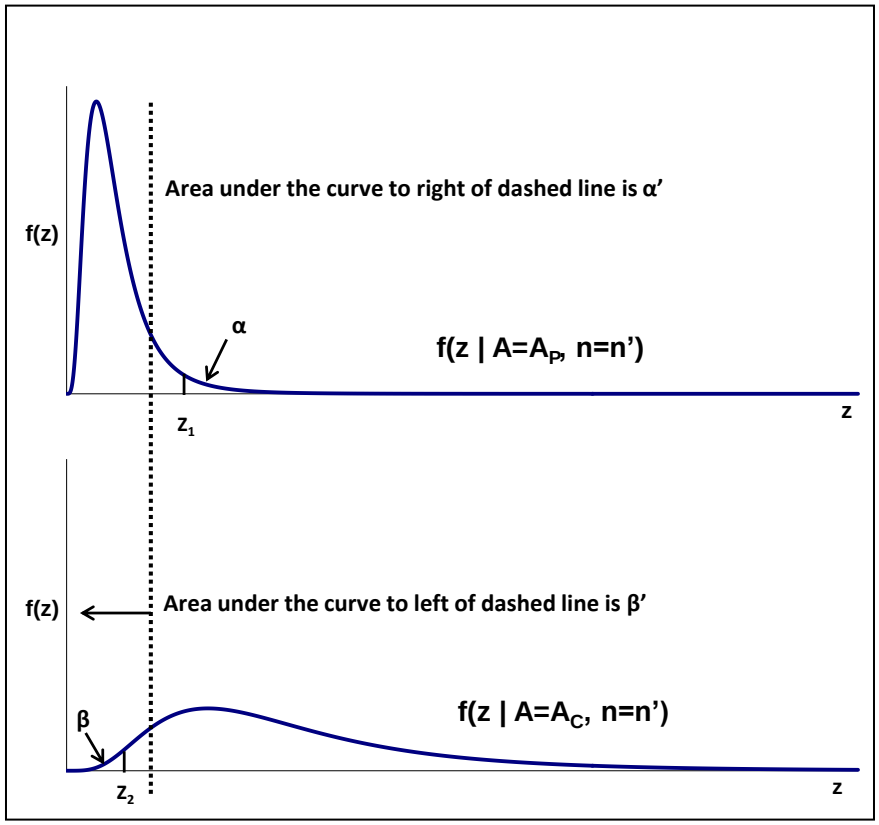

Figure 13. Graphical Representation of Sub-Optimal FNF Test Plan and non-compliant Producer and Consumer Risk Levels under Time-Constrained n'

The area under the curve to the left of any point z is defined as

$$f_1(z \mid n,\ A) = \int_0^z \frac{\Gamma(n^2)}{[\Gamma(n)]^2}\left(\frac{A}{1-A}\right)\frac{x^{n-1}}{(1+[A/(1-A)]x)^2}\,dx \qquad (37).$$

Based on time constraint and the general definition of producer and consumer risks,

$$f_1(z \mid n',\ A = A_P) = (1-\alpha') \qquad (38)$$

$$f_1(z \mid n',\ A = A_C) = \beta' \qquad (39)$$

When applying the unconstrained FNF strategy, A_P, A_C, α, and β are specified and equations (32) and (33) are solved for optimum values n^* and z^*. Once a time constraint is placed on the test plan, A_P, A_C, and $n=n'$ are specified and equations (38) and(39) must be solved for α', β', and z'. Three unknown variables in two equations create a pair of dependent equations relating α' and β' as a function of z'.

The relationship between α, β, and z is shown graphically in Figure 13, under the conditions of optimality or $n^* = 6$. Since α and β are a function of z, their relationship can be shown directly on a plot with z on the horizontal axis, and individual curves for α and β. The former curve is the complement of the cumulative density function representing the probability of rejection given $A=A_P$ and n. The latter curve is the CDF representing the probability of acceptance given $A=A_C$ and n. In this case, $n=n^*$

The example presented in Figure 14 represents the Chapter 3 scenario with A_P = 0.95, A_C = 0.90, α = 0.10, and β = 0.10. The graph of α and β is shown as a function of z for these parameters and the optimal number of cycles, n^* = 6. A vertical line perpendicular through any point on the z axis identifies the α and β values associated with the producer and consumer risk for a test plan of (6, z^*). Note at the point z = 0.113, α and β are very close to 0.10 as stated in the scenario. Any z associated with α > 0.10, or β > 0.10 is not an optimal solution for the unconstrained problem.

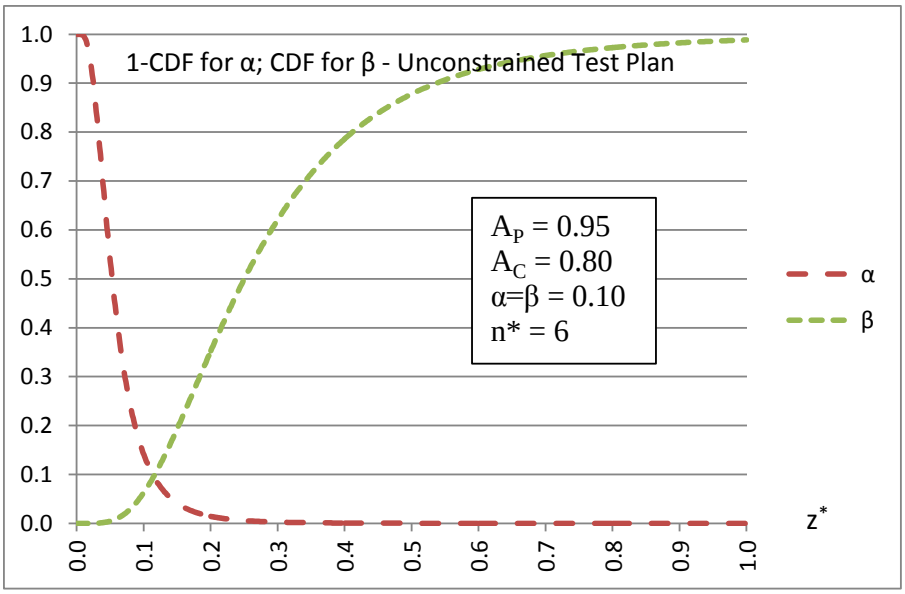

Figure 14. CDF for Probability of Rejection when $A=A_P$ (α) and Probability of Acceptance when $A=A_C$ (β) for Fixed Number of Failure ADT Strategy

The graphical representation for a time-constrained strategy is shown in Figure 15. The relationship is plotted by numerically solving (38) and (39) for increments of *z'* from 0 to 1 for a specified $n = n'$, where test duration for FNF is established in Equation (36) as a direct function of *n'*. While *z* can be greater than 1, such a value implies that downtime is greater than uptime for the system. This is rarely practical in actual system operation; therefore, the upper bound of 1 is considered reasonable for *z*.

The example presented represents a time constraint placed on the Chapter 3 scenario with $A_P = 0.95$, $A_C = 0.90$, $\alpha = 0.10$, and $\beta = 0.10$. Under time constraint, the decision maker specifies a maximum test time and equation (36) defines the number of allowed cycles. Assuming the example scenario values, and $n' = 2$, the new relationships between α, β, and *z* are plotted in Figure 15. Note that the curves for both α and β have shifted upward. The move from 6 cycles to 2 cycles has shifted the point of equal risk from 0.10 to 0.235.

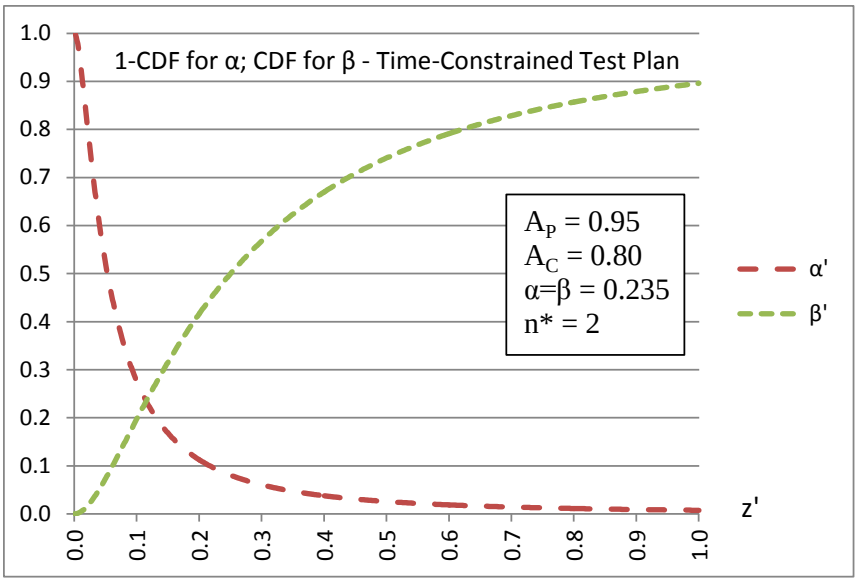

Figure 15. CDF for Probability of Rejection when $A=A_P$ (α') and Probability of Acceptance when $A=A_C$ (β') for Time-Constrained FNF ADT Strategy

The impact of placing time constraint on the FNF strategy is the simultaneous increase in producer and consumer risk levels, holding the associated availability levels constant. Test duration is specified in terms of *n'*, a sub-optimal number of failure-repair cycles with respect to the original producer and consumer risk statements. For each *n'*, an integer, there is a unique pair of relationships between the producer and consumer risks and the critical test metric *z'*. These relationships can be stated as a set of triples (α', β', z') where each triple is a solution to the pair of equations (38) and (39) for a given value of *n'*. These triples represent the points on the x-axis, the α-plot, and the β-plot for a vertical line drawn through the graph, and define all alternative test plans under *n'*.

4.2.2 Time Constraint on the Fixed Test Time Test Strategy

The fixed test time strategy uses the statement of the producer and consumer risk to specify the parameters A_P, A_C, α, and β. Additionally, the mean of the failure distribution under the null hypothesis, m_u, is stated from specified requirements for the system, or derived from historical data. From these parameters, U_P, U_C, $\lambda_{1-\alpha}$, $\lambda_{1-\beta}$, and D

are calculated. The optimal test plan (T^*, $A(T)^*$) defines the duration of the test, T^*, and the critical test value $A(T)^*$. These values are calculated directly from equations (24) and (25).

Under time constraint, *T'* is defined as the maximum time available for testing and is less than T^* by definition. The availability levels defined by the producer and consumer risk statements, A_P, A_C respectively, are held constant. Substituting *T'* into equation (24) for T^* yields a single equation in two unknowns. This forms a dependent relationship between α' and β'. Solving equation (24) for the term $\lambda_{1-\alpha'}$ the dependency between α' and β' is defined as:

$$\lambda_{1-\alpha'} = \sqrt{\frac{T'}{2m_u}}\frac{(D-1)}{\sqrt{1-U_P}} - \lambda_{1-\beta'}\frac{(1-DU_P)\sqrt{D}}{(1-U_P)} \tag{40}$$

or alternatively, solving for $\lambda_{1-\beta'}$ yields

$$\lambda_{1-\beta'} = \sqrt{\frac{T'}{2m_u}}\frac{(D-1)}{\sqrt{D}}\frac{\sqrt{1-U_P}}{(1-DU_P)} - \lambda_{1-\alpha'}\frac{(1-U_P)}{(1-DU_P)\sqrt{D}} \tag{41}$$

Under the normal assumption of the FTT methodology, $\lambda_{1-\alpha'}$ and $\lambda_{1-\beta'}$ are standard normal random variables. Every pair of (α', β') that solves either (40) or (41) can be substituted into equation (25) to obtain the associated critical test value A(T)'.

Consider the scenario from Chapter 3 with the following risk statements

- Producer Risk Statement: A system with a true availability of 0.95 will pass the test with probability 90%.
- Consumer Risk Statement: A system with a true availability of 0.80 will pass the test with probability 10%.

and assume a mean up time of 1900 hours. The optimum test strategy was calculated from (24) and (25) to be (4746.5, 0.8941). The decision maker has limited the test

duration to 3500 hours. Setting T'=3500 and solving equation (40) for β' from 0.001 to 0.999 in increments of 0.001 yields Figure 16.

A vertical line at any point in the graph identifies the dependent α' and β' values as well as the associated A(T)' critical value. Notice that the point of equivalent producer and consumer risk has moved from 0.10 to 0.135. Each triple (α', β', A(T)') specifies an alternative test plan under T'.

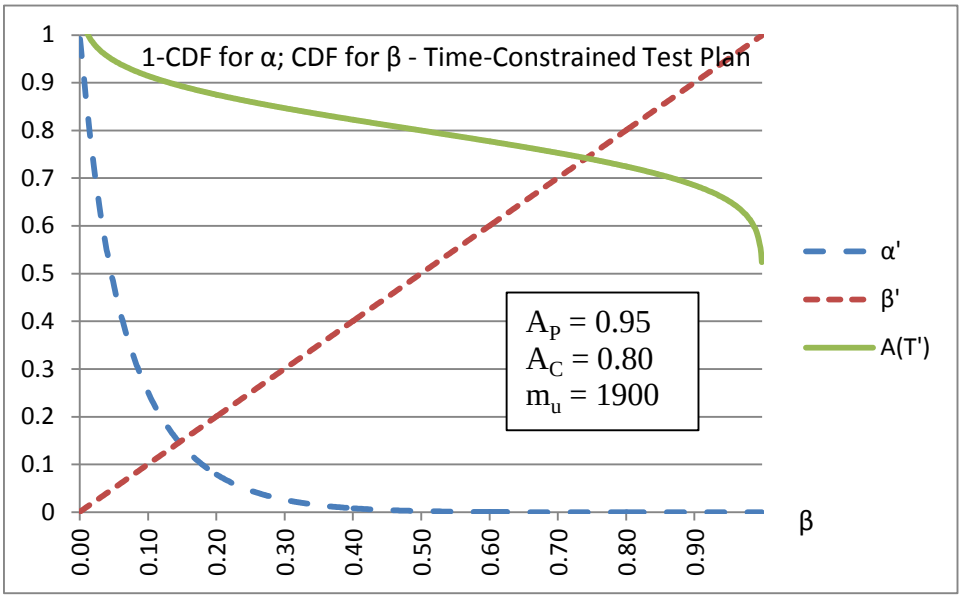

Figure 16. Plot of α' and A(T)' as a function of β' for time-constrained Fixed Test Time ADT Strategy (A_P = 0.95, A_C = 0.80, T' = 3500 hours)

4.2.3 Challenge of ADT under Time Constraint

The ability to define valid test plans under a decision maker specified time constraint has been demonstrated for both FNF and FTT strategies. In each case, a user defined time constraint has been converted into the triple (α', β', Crit), where Crit is z' for FNF and A(T)' for FTT strategies. This triple defines the set of test plans that meet the decision maker defined levels for specified A_P, A_C, and time constraint.

For example, test plans are considered for FTT under the assumptions $A_p = 0.95$, $A_C = 0.80$, and T' = 3500 hours. This set of test plans is defined by vertical lines through β values in Figure 16. Three options include test plans defined by the triples (0.135, 0.135, 0.8941), (0.100, 0.160, 0.8849), and (0.213, 0.100, 0.9096). The first test plan represents an equal producer and consumer risk at 0.135. The second test plan holds the producer risk at its original level of 0.10, and solves for the associated consumer risk and critical values of 0.160 and 0.8849 respectively. The third alternative holds the consumer risk at the original level of 0.10, and solves for the associated producer risk and critical test values of 0.213 and 0.9096 respectively.

Alternative test plans such as those discussed can be generated under a specific time constraint, but there is no understanding of the relative goodness of one alternative to another. A best test plan cannot be designated without relative value between the alternatives. Some contextual basis of comparison must be used to establish relative value of a test plan and ultimately decide on the best plan within the set of options.

This research has consistently recognized the concern for time and cost throughout the discussion. Time is the stated constraint for the test planning. As such, cost becomes the most reasonable context within which to evaluate alternative plans and provide meaningful insight to the decision maker.

4.3 Cost of Availability Risk Mitigation

Across the lifecycle of a system, trade-off decisions are constantly made to optimize some critical aspect of the system subject to competing objectives. In many if not all of these cases, cost is a critical element in the basis of decision making. Referring back to the Availability-Cost relationship in Figure 7, outside of some optimum value the

cost increases whether we lower availability or increase it. A similar effect will be shown to prevail for the risk mitigation portion of availability cost. Producer and consumer risk must be balanced at an optimal point. Outside of this point, expected cost will increases regardless of the direction of change.

Equation (19) in section 2.4 provides a high level calculation for life cycle cost. Within that model, the cost elements for validation and warranty have the strongest relationship to risk mitigation. Validation testing is the methodology used to verify a system complies with an expected level of performance. This testing is statistical in nature with some non-zero chance of error, and will not be absolute unless 100% of the population is tested for 100% of the service life. The longer the duration of test, the greater the certainty the system meets its required performance level and the higher the cost. This statement defines the competing objectives; certainty verses the cost of an incorrect decision.

An equation for the cost of availability risk mitigation is developed. This equation will provide a framework to evaluate producer and consumer risk levels defined by individual test plans within a common time-constrained test strategy. Elements within the cost of availability risk mitigation equation include:

C_{ARM} =	Cost of Availability Risk Mitigation
C_V =	Cost of Validation for Availability including all overhead burden
$[C_V]_{Fixed}$ =	Non-recurring fixed element of Cost of Validation for Availability
$[C_V]_{Recur}$ =	Recurring element of Cost of Validation for Availability
C_W =	Cost of Warranty

$[C_W]_{Fixed}$ = Fixed expense under which some party accepts cost of unavailability above a specified level of performance (may have limitation or be unlimited)

$[C_W]_{Recur}$ = Recurring element due to probabilistic exposure to consumer for increased sustainment cost for system that does not meet performance requirement over and above that covered by $[C_W]_{Fixed}$

Furthermore, define the following relationships between these cost elements:

$$C_{ARM} = C_V + C_W \tag{42}$$

$$C_V = [C_V]_{Fixed} + [C_V]_{Recur} \tag{43}$$

$$C_W = [C_W]_{Fixed} + [C_W]_{Recur} \tag{44}$$

and by substituting (43) and (44) into (42),

$$C_{ARM} = [C_V]_{Fixed} + [C_V]_{Recur} + [C_W]_{Fixed} + [C_W]_{Recur} \tag{45}.$$

Equation (45) represents the validation cost and the cost of risk exposure for system availability. One could argue that warranty cost is not risk exposure. However, once purchased, the consumer may never realize the value of the cost. If the cost of sustainment covered under the warranty is less than the cost of the warranty, the consumer loses money. Additionally, warranty provides some negative risk exposure since the fixed cost will limit the expense to the consumer.

When performing ADT, there are two inherent risks that have been previously defined. The first is the producer's risk α, or the probability of rejecting a system that meets the stated requirements. The second is the consumer's risk β, or the probability of accepting a system that does not meet the stated requirements. Both errors have an impact on the C_{ARM} defined by equation (45).

If a system which meets the stated requirements is rejected, further testing is required. While the fixed, non-recurring costs will not change, the recurring cost

associated with reevaluating the system must be expended. In fact, every time ADT is performed on the system, the probability of a false rejection is possible. Thus, the true cost of validation for availability is

$$C_V = [C_V]_{Fixed} + \{[C_V]_{Recur} + \alpha[C_V]_{Recur} + \alpha^2[C_V]_{Recur} + \alpha^3[C_V]_{Recur} + \cdots\}$$

where α^n indicates the probability of failing *n* consecutive tests when the system meets design requirements. Simplification of the geometric series in this equation yields the Cost of Validation for Availability

$$C_V = [C_V]_{Fixed} + \left(\frac{1}{1-\alpha}\right)[C_V]_{Recur} \quad for\ |\alpha| < 1 \tag{46}$$

Since $0 \leq \alpha \leq 1$ by definition, then equation (46) holds as long as $\alpha \neq 0$ and $\alpha \neq 1$. These values for α would never be used in a practical application, therefore, the restriction has no impact on this research.

If the consumer risk is realized, then a system that does not meet stated requirements is accepted and fielded. If this occurs, the fixed cost of warranty will be incurred regardless of the test conditions. However, the additional cost of sustainment due to lower availability becomes a function of the probability of accepting a system with $A=A_C$. Therefore, the true cost of warranty is the fixed cost plus the product of β and $[C_W]_{Recur}$. This relationship is stated as:

$$C_W = [C_W]_{Fixed} + \beta[C_W]_{Recur} \tag{47}$$

and substituting (46) and (47) into equation (42) yields

$$C_{ARM} = [C_V]_{Fixed} + \left(\frac{1}{1-\alpha}\right)[C_V]_{Recur} + [C_W]_{Fixed} + \beta[C_W]_{Recur} \tag{48}$$

4.4 Model for Cost of Availability Risk Mitigation under Time Constraint

The model for the cost of availability risk mitigation given a time constraint on ADT is very straightforward for both FNF and FTT strategies. Section 4.2.1 defines the

triple (α', β', z') for test plans under a time constrained FNF strategy. Furthermore, section 4.2.2 defines the equivalent triple (α', β', A(T)') for test plans under a time constrained FTT strategy. In both of these cases, unique α'and β' values are linked to the critical value for the test metric. Therefore, a single evaluation of equation (48) exists for each discrete value of the critical test metric for FNF and FTT strategies.

The four cost elements within equation (48) are constant across all values of α and β. Restated, the fixed and recurring cost of test, warranty, and sustainment are independent of the producer and consumer risk level. For example, if the consumer risk moves from 0.10 to 0.20, the sustainment cost for a system with $A=A_C$ does not change. The risk exposure to the decision maker changes, but not the sustainment cost itself.

The objective for establishing a model for C_{ARM} will be to substitute the risk terms in equation (48) with equivalent mathematical statements, equations (38), (39), and (40), that allow direct evaluation of the expected cost. Ideally, these statements are in terms of our test parameters or a derivative of them. For FNF, equations (38) and (39) provide these equivalent calculations for α, and β. Rearranging the terms in (38) and reversing (39) define the following equalities:

$$\alpha' = 1 - f_1(z' \mid n', A_P) \tag{49}$$

$$\beta' = f_1(z' \mid n', A_C) \tag{50}$$

Substituting (49) and (50) into equation (48) creates the following C_{ARM} equations for FNF:

$$C_{ARM} = [C_V]_{Fixed} + \left(\frac{1}{f_1(z' \mid n', A_P)}\right)[C_V]_{Recur} + [C_W]_{Fixed} + f_1(z' \mid n', A_C)[C_W]_{Recur} \tag{51}$$

This cost equation is evaluated over the range of $0 < z' < 1$ to determine the least C_{ARM} as a function of z'.

For the FTT strategy, specification of the C_{ARM} equation begins with equation (40). Taking the CDF function Φ for the standard normal distribution of both sides and solving for α yields

$$\alpha' = 1 - \Phi\left(\sqrt{\frac{T'}{2m_u}}\frac{(D-1)}{\sqrt{1-U_P}} - \lambda_{1-\beta'}\frac{(1-DU_P)\sqrt{D}}{(1-U_P)}\right) \tag{52}$$

over the range of $0 < \beta' < 1$. Substituting equation (52) into equation (48) results in

$$C_{ARM} = [C_V]_{Fixed} + \left(\frac{1}{\Phi\left[\sqrt{\frac{T'}{2m_u}}\frac{(D-1)}{\sqrt{1-U_P}} - \lambda_{1-\beta'}\frac{(1-DU_P)\sqrt{D}}{(1-U_P)}\right]}\right)[C_V]_{Recur}$$

$$+[C_W]_{Fixed} + \beta'[C_W]_{\text{Recur}} \tag{53}$$

Equations (51) and (53) can be evaluated over the stated ranges of z' and β using a very simple spreadsheet. By evaluating C_{ARM} values over the range of 0 to 1 using increments as large as 0.001, the cost function can be generated and graphed. The minimum value of C_{ARM} across the range will define the triple representing the specific test plan that provides the least expected cost of risk mitigation. A process, supporting the decision maker in selection of a best test plan, based on this concept will be presented in Chapter 5.

4.5 Conclusions

When faced with schedule slippage and cost overruns, decision makers are under pressure to reduce the test duration of validation testing to save time and money. When holding the availability levels defined in the producer and consumer risk statements steady, a reduction in test time results in a simultaneous increase in α and β. The magnitude of this change is unknown to the decision maker.

An approach was developed that defines alternative test plans under time-constrained testing for both FNF and FTT strategies. These alternative test plans are characterized by the triples (α', β', z') for FNF and (α', β', A(T)') for FTT. In both cases, a unique pair of risk variables is linked to a specific value of the critical test metric. However, uncertainty exists concerning the relative value of competing test plans under a specified time constraint.

A valid context for evaluation was established with the cost for availability risk mitigation (C_{ARM}). A cost equation was developed including fixed and variable cost elements. The type I and type II errors for hypothesis testing under ADT were integrated into the C_{ARM} equation. This integrated cost equation provides the basis to compare alternative test plans under the time-constrained FNF strategy and determine the least cost alternative. This basis was valid for the time constrained FTT strategy as well.

Equations used to calculate the risk values for alternative test plans were solved for α and β, and then substituted into the C_{ARM} equation. By direct evaluation of the C_{ARM} equation over the range of $0 < z' < 1$ for the FNF strategy, and over the range of $0 < \beta < 1$ for the FTT strategy, the triple representing the least cost test plan under the specified time constraint is identified. This provides the program manager with a decision support tool to aide in balancing risk levels to minimize C_{ARM} in the event of time constraint to the test and evaluation program.

4.6 Future Research

Future research opportunities exist to improve the model and to evaluate the performance of FNF verses FTT strategies. With respect to the C_{ARM} model, there are two opportunities for improvement. These include the inclusion of a cost for time delay

in fielding, and consideration for a conditional probability on the value of a fixed price warranty.

The C_{ARM} model considers the fixed and variable cost of validation testing. The variable cost measures the additional fiscal cost of testing when a system meeting user requirements was rejected. There is an additional opportunity cost that was not deliberated. If a system meeting specified availability requirements is rejected and subject to retest, a time delay in fielding the system is created. There is an opportunity cost associated with this time delay that should be evaluated. In the case of a new capability that doesn't exist, it is quite possible that a less than perfect solution now is more valuable than a perfect solution in the future.

The second model consideration involves the impact of a fixed price warranty on the incremental sustainment cost associated with fielding a system that does not meet user requirements. Consider the example used throughout Chapter 4. If $A_P = 0.95$ and $A_C = 0.80$, there is the possibility that the consumer could purchase a fixed price warranty that covers the cost of sustainment for availability less than 0.95 but limited at A_{True} equal to 0.82. This would imply that the consumer's risk exposure is limited to a lower value. This change in exposure could lead to a different optimal test plan. The conditional probability of the value of $[C_W]_{Recurr}$ given the cost of the fixed price warranty could be integrated into the C_{ARM} model.

The final consideration for future research is the comparative evaluation of the approaches to time constrained FNF and FTT. It is of value to a decision maker to understand which approach, if any, provides a better or more robust recommendation for ADT planning. This approach might follow the method taken in Chapter 3.

CHAPTER 5

A PROCESS FOR MANAGING RISK LEVELS IN AVAILABILITY DEMONSTRATION TESTING UNDER TIME CONSTRAINT

5.1 Introduction

Management of a complex system from concept to disposal is challenging. Software applications have been developed to help engineering teams understand and manage the definition, allocation, and dependency of requirements across functional and physical baselines within the system engineering process. Aides such as these enable a decision maker to understand the multifaceted relationships that exist and to properly respond to the impact of unanticipated change that can ripple throughout the process. In the absence of a decision support process for time constrained ADT, the decision maker is ignorant to change in risk levels that result from dependencies and thus unable to manage the overall risk exposure of the program. The concepts and equations developed in Chapters 3 and 4 provide the basis for a decision support process to allow a decision maker to balance risk in time-constrained ADT.

Availability demonstration testing has been defined as a mitigation tool used to minimize the risk of fielding a non-compliant system with reduced mission success and increased life cycle cost. While three test strategies have been identified, nothing within the available literature suggested the comparative strengths and weaknesses between these approaches. Chapter 3 of this research provides evidence of the performance of

each strategy with respect to quality of decision, and timeliness. This information has great utility for the decision maker when selecting the strategy to adopt in development of a program's ADT plan.

Timing of compliance testing such as ADT is late in the acquisition lifecycle, usually after production has started. Schedule and budget issues as well as the need for timely delivery of the system to the end user can create pressure on the program manager to reduce compliance test duration. Chapter 4 has shown that such a reduction in test time while holding producer and consumer availability targets constant will result in a simultaneous increase in α and β risk levels.

The increase in risk levels is not intuitive, and the decision maker can be left unaware of this change to risk exposure. Section 4.2 provides an approach for understanding the change in risk due to a time constraint on testing, defined as a set of test plans with unique and specific α and β risk levels associated with each critical test value. Furthermore, section 4.3 defines an equation for the cost of availability risk mitigation which provides a basis for comparison of alternative test plans under a specified time constraint. Consideration of the impact of type I and type II errors in demonstration testing provides for the direct evaluation of the C_{ARM} for each alternative test plan. Thus, the test plan that minimizes this risk mitigation cost defines the preferred values for producer and consumer risk.

The integration of these concepts into a single process will provide a decision maker with the robust decision support needed to properly understand risk exposure and select the most desirable test plan when facing a time constraint. This chapter will define a recursive process that enables the decision maker to apply a time constraint to an

optimized ADT, and identify the plan from the set of alternatives generated under the restriction which has the lowest C_{ARM}. This process will then be demonstrated with an example that utilizes a simple spreadsheet to perform calculations and optimization.

5.2 Time-constrained Availability Demonstration Test Risk Specification Process

The decision maker requires a support tool to manage producer and consumer risk levels during the development of an availability demonstration test plan. Such a tool will be specified using the knowledge presented in prior chapters of this research using a two stage approach. The flow for this process is shown in Figure 17.

In the initial stage, the producer and consumer risk statements are generated specifying target availability and risk levels. Afterward, a test strategy will be selected based on the program priority with respect to quality of decision and timeliness of testing. Quality of decision represents the precision of a test plan's true producer and consumer risk levels, and timeliness is measured by the expected duration of the demonstration test.

The second stage of the process will focus on development of the exact test plan. This stage follows one of two branches based on the test strategy selected in the first stage. Using the availability and risk levels specified in the producer and consumer risk statements, an optimal test strategy is generated. If the time duration of this strategy does not meet the decision maker's needs, a time constraint will be specified for the test event. Given the specified time constraint, the set of alternative test plans is generated. These test plans are then evaluated using the C_{ARM} equation to determine the minimum cost test plan. The decision maker evaluates the risk levels associated with the plan and can decide to accept the risks as defined, or redefine the time constraint. If the latter choice is

selected, then a new set of test plans under the revised time constraint is generated and the process continues.

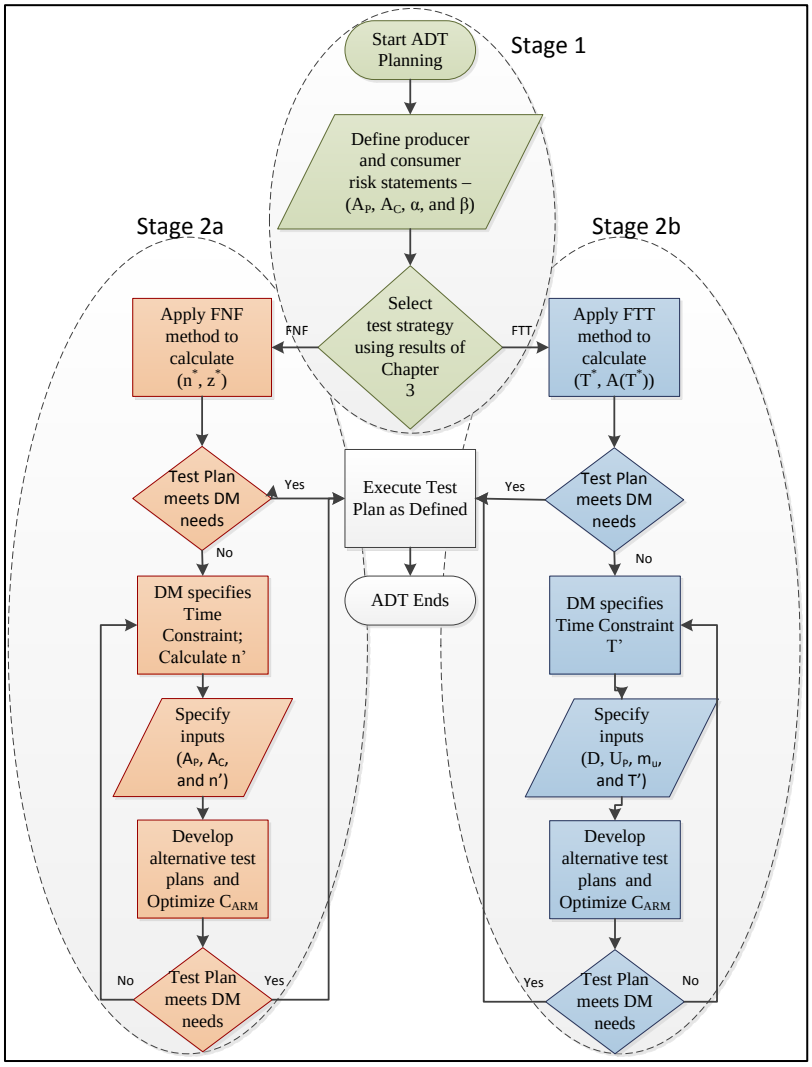

Figure 17. Time-Constrained Availability Demonstration Test Risk Specification Process Flow Chart

Implementation of the equations and methods developed in prior chapters and applied in this process can be accomplished very easily using a simple spreadsheet. Examples presented in this chapter were generated in a four tab Excel© spreadsheet. The content of the spreadsheet is not a significant result of the research. This reference is

provided to convey that special software or programming expertise is not required to implement the process.

5.2.1 Stage 1 - Selection of Test Strategy

The first step in the time-constrained ADT risk specification process is development of producer and consumer risk statements. These statements result in the specification of A_P, A_C, α and β. The parameters from the risk statements will be the same regardless of the test strategy selected. As discussed in Chapter 4, only FNF and FTT test strategies are considered under a time constraint. The most desirable test strategy will depend upon the decision maker's preference for decision quality or timeliness.

The specification of a test plan under either strategy is dependent on the producer and consumer risk levels. However, the actual probability of accepting a good system or rejecting a bad system will vary from these target levels. The magnitude of deviation of the actual risk level from the target risk level for both producer and consumer risk is the measure of decision quality. If a decision maker is sensitive to these values, then they are quality centric.

The actual duration of a test is the measure of timeliness. A test plan that produces a decision faster is considered timelier than slower alternatives. Hence, if the decision maker's priority is getting a result fastest, then they are time centric.

If the decision maker is quality centric and places a priority on a test plan that produces risk levels closest to the target values, then the preferred strategy is FNF according to the analysis of variance results in Chapter 3. If the decision maker places a priority on timeliness over quality of decision, then Chapter 3 concluded that FTT was

preferred. It is noted that two approaches to FTT, pre-truncation and post-truncation, were discussed. These methods were both preferred over the FNF strategy when considering timeliness of testing. Within the two choices of FTT, pre-truncation is timelier than post-truncation while post-truncation has better quality of decision than pre-truncation.

Discussion throughout the remainder of the sub-sections in section 5.2 will utilize the following risk statements:

- Producer Risk: A system with a true availability of 0.95 will pass the test 90% of the time
- Consumer Risk: A system with a true availability of 0.80 will pass the test 10% of the time

Given these statements, then $A_P = 0.95$, $A_C = 0.80$, $\alpha = 0.10$ and $\beta = 0.10$. It is also assumed that the mean uptime, m_u, is equal to 1900 hours. This is equivalent to the example scenario used in previous chapters.

5.2.2 Stage 2a – Minimum C_{ARM} Test Plan – Fixed Number of Failures Strategy

If the decision maker is quality centric, the process will follow the branch to stage 2a in Figure 17. This path will utilize the FNF strategy for ADT. The first step in this path generates an optimum FNF test plan utilizing equations (32) and (33). The solution method described in section 3.1 yields an optimal test plan of (6, 0.113). This test plan has an expected duration of 12,000 hours based on equation (34).

If the decision maker is satisfied with this plan, testing is implemented as planned. Otherwise, the decision maker must specify a time constraint on the test. Within this example, a time constraint of 6,000 hours is used. Based on equation (36), this equates to $n' = 3$. Utilizing equations (38) and (39), the set of alternative test plans are generated for

$0 < z' < 1.0$. This set of test plans is graphically shown in Figure 18. A vertical line through the z' axis defines each test plan as the triple (α', β', z').

Once the set of test plans are defined, their associated C_{ARM} value is calculated utilizing equation (51). Table 15 shows the cost elements assumed for this example.

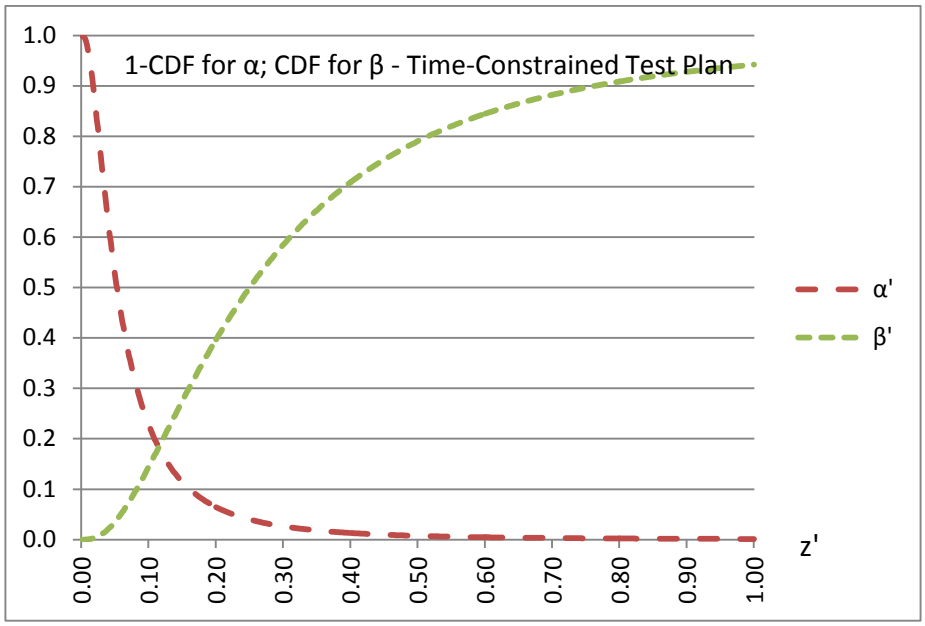

Figure 18. Alternative FNF Test Plans for $A_P = 0.95$, $A_C = 0.80$, and n' = 3

Table 15

Elements for Cost of Availability Risk Mitigation for Example Scenario

$(C_V)_{Fixed}$	25000
$(C_V)_{Recur}$	15000
$(C_W)_{Fixed}$	50000
$(C_W)_{Recur}$	75000

The resulting cost values are summarized in Figure 19. The minimum C_{ARM} value occurs for the test plan (0.322, 0.091, 0.078) with a cost of 103970. Under this test plan, the revised risk statements are:

- Producer Risk: A system with a true availability of 0.95 will pass the test 67.8% of the time
- Consumer Risk: A system with a true availability of 0.80 will pass the test 9.1% of the time

It is up to the decision maker to accept or reject these risk levels. If accepted, the test plan is implemented. If the risk levels are not acceptable, then additional time must be allowed for the test. In this event, the process will loop back to the specification of time constraint step, and the process continues.

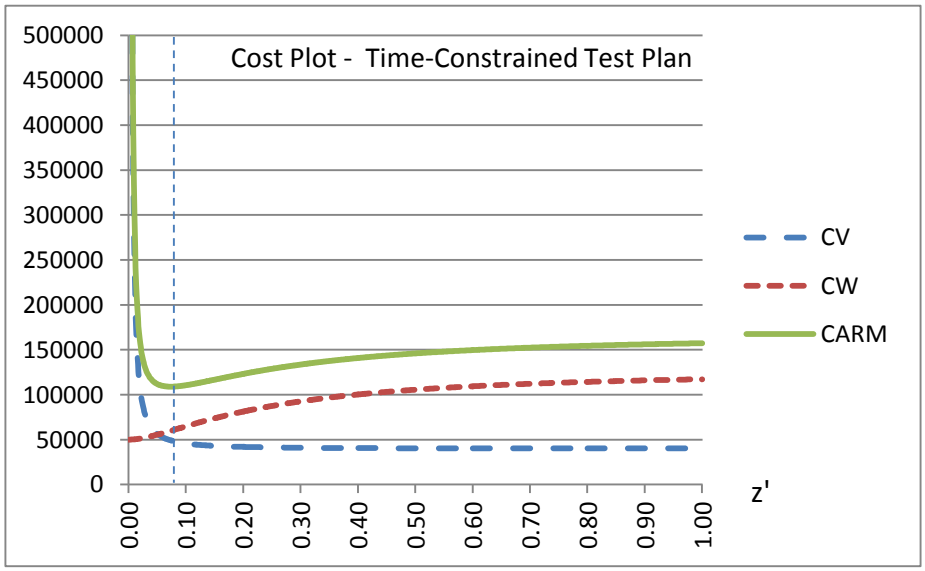

Figure 19. C_{ARM} for alternative FNF Test Plans with $A_P = 0.95$, $A_C = 0.80$, and n' = 3 and Cost elements $[C_V]_{Fixed} = 25K$, $[C_V]_{Recur} = 15K$, $[C_W]_{Fixed} = 50K$, $[C_W]_{Recur} = 75K$

A noteworthy observation can be made concerning the C_{ARM} curve and its components. Both components, C_V and C_W, converge to a fixed value as z' approaches 1. The cost of verification is very high for z' close to zero, which equates to α values close to 1. Since (1 - α) is in the denominator of the coefficient for the recurring cost, α close to 1 drives this term to infinity. As z' increases and α decreases, the cost of verification approaches the sum of fixed cost and the variable cost.

Conversely, the consumer risk β and z' vary directly, so as z' increases, β increases. Therefore, the cost of warranty is equal to the fixed cost at z' = 0, and approaches the sum of the fixed cost and variable cost as z' approaches 1. These observations will be valid for both test strategies discussed in sections 5.2.3 and 5.2.3.

5.2.3 Stage 2b – Minimum C_{ARM} Test Plan – Fixed Test Time Strategy

Decision makers who place priority on timely completion of ADT will pursue the branch to stage 2b in Figure 17. This path utilizes the FTT strategy for ADT. The first step in this path requires the generation of an optimum FTT test plan utilizing equations (24) and (25). The solution for the optimal test plan for the example scenario is (4746.5, 0.8941). This test plan has an expected duration of $T^* = 4746.5$. The decision maker has placed a time constraint of 3200 hours on the test.

Given the derived values for D and U_P, and the defined values for T' and m_u, equation (40) is solved for $0 < \beta < 1$ to generate the set of alternate test plans under the time constraint T' = 3200. These plans are shown in Figure 20.

Using the cost elements specified in Table 15, equation (53) is solved to obtain the C_{ARM} values for the alternative test plans. The validation and warranty cost curves, as well as the composite C_{ARM} curve are shown in Figure 21. Note that A(T)' is monotonically decreasing in Figure 20, while it is increasing in Figure 21. Therefore, the references to α and β are reversed between the two graphs.

The minimum C_{ARM} value occurs for the test plan (0.241, 0.104, 0.9126) with a cost of 102556. Under this test plan, the revised risk statements are:

- Producer Risk: A system with a true availability of 0.95 will pass the test 75.9% of the time

- Consumer Risk: A system with a true availability of 0.80 will pass the test 10.4% of the time

If the decision maker is satisfied with these risk statements, the test plan is implemented. If the risks are too high, then the process loops back to the determination of the time constraint and a new T' greater than the previous value must be selected. The process proceeds until acceptable values for α', β', and T' are defined.

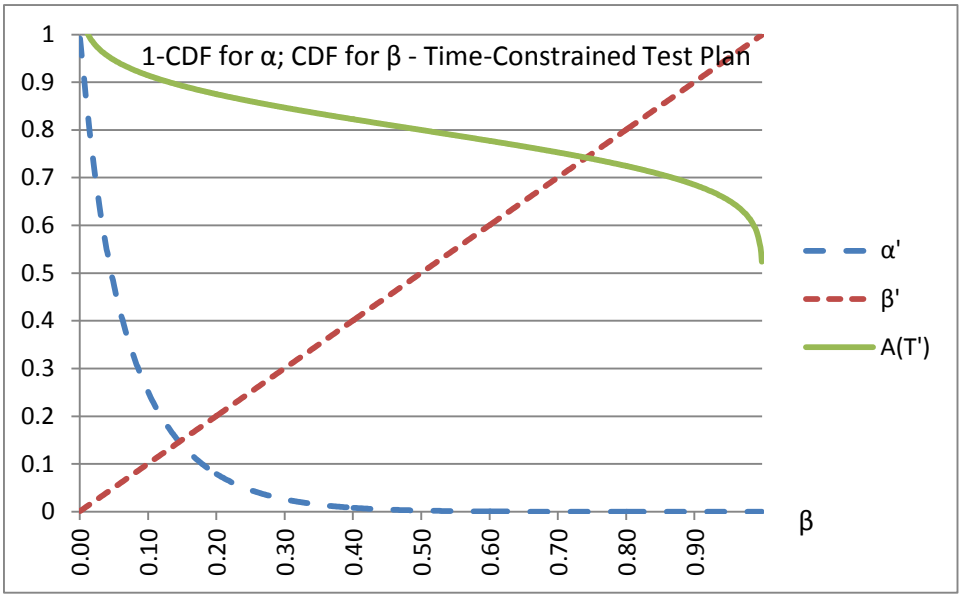

Figure 20. Alternative FTT Test Plans for D = 4, U_P = 0.05, m_u = 1900, and T' = 3200

5.3 Case Study

The surface navy electronic system ADT design case study presented in this section represents the application of the proposed methodology developed within the research. The analyzed case study deals with a complex radio frequency electronics system that is managed through spiral development applied over a long life cycle of thirty or more years. Each spiral upgrade targets one or more specific subsystems to improve technical performance and supportability of the overall system. Data was provided by development team members responsible for RMA engineering and testing of the system.

The actual values used in the example have been modified to protect the sensitive nature of the data. This assessment should help the reader understand the practical application of this approach. Both strategies will be presented for clarity sake, although in a real case, the decision maker would choose one or the other based on program priorities as discussed in section 5.2.1.

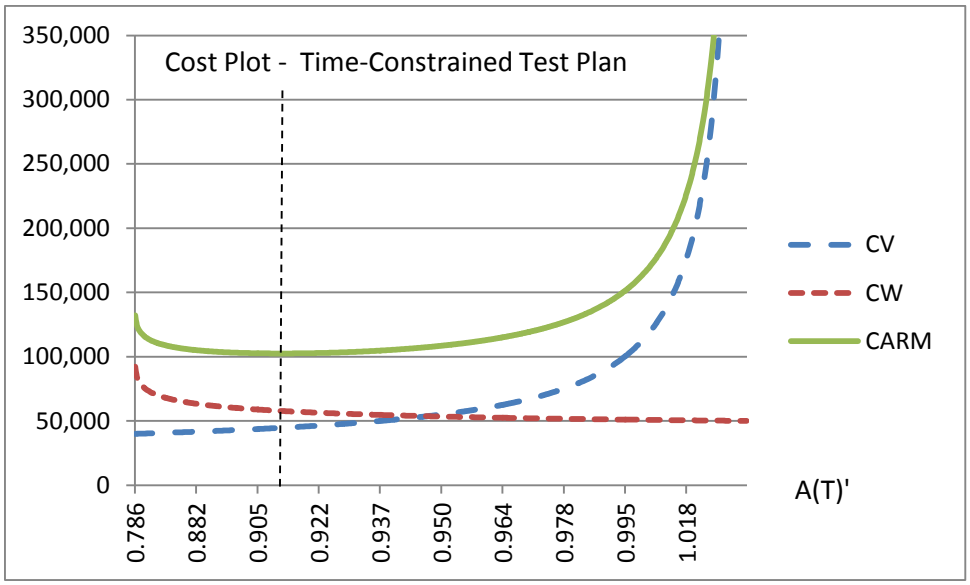

Figure 21. C_{ARM} for alternative FTT Test Plans with D = 4, $U_P = 0.05$, $m_u = 1900$, and T' = 3200; Cost elements $[C_V]_{Fixed} = 25K$, $[C_V]_{Recur} = 15K$, $[C_W]_{Fixed} = 50K$, $[C_W]_{Recur} = 75K$

The goals of the ADT for the ongoing design of the current spiral are expressed within the producer and consumer risk statements:

- Producer Risk: A system with a true availability of 0.90 will pass the test 90% of the time
- Consumer Risk: A system with a true availability of 0.80 will pass the test 10% of the time

Based on the risk statements, $A_P = 0.90$, $A_C = 0.80$, $\alpha = 0.10$, and $\beta = 0.10$. From these values, D = 2 and $U_P = 0.10$ are derived. Using historical data augmented with engineering analysis to account for proposed design changes, the system is estimated to have a mean uptime or MTBF equal to 1806 hours. Under the null hypothesis of A =

0.90, the mean downtime or MTTR is estimated at 201 hours. The decision maker over this program has defined a maximum allowable test time for ADT of 8000 hours. This will be accomplished by running two initial production units simultaneously for 4000 hours each.

Verification test costs for this case have been estimated and are summarized in Table 16. These costs are grouped by fixed and recurring, and are specified in accordance with the elements presented in Table 5. Those elements with zero values are included in the table for completeness.

This program does not purchase a vendor or third party warranty. As such, $[C_W]_{Fixed} = 0$. The recurring element of the cost of warranty is the estimate of the additional sustainment cost to support a system with $A = A_C$ over the sustainment cost estimated for $A = A_P$. Using data from a previous spiral, the average cost per maintenance event was calculated as the sum of material and the cost of labor.

The challenge in calculating the delta sustainment cost lies in the fact there are several ways A_C can vary from the definition of A_P. The MTBF for A_P can remain constant while increasing the MTTR. The MTTR for A_P can remain constant while the MTBF is reduced. Or, MTBF can be reduced simultaneously with an increase in MTTR. Each of these three conditions was evaluated. The delta sustainment cost for holding MTBF constant and increasing MTTR was the lowest. The delta cost for holding MTTR constant and lowering MTBF was the highest. The third option was evaluated by lowering MTBF by the same percentage as the increase in MTTR. This approach resulted in a delta sustainment cost that was higher than, but close to, the mean of the first two options.

This case study utilized the third approach since it provided a middle ground for the variable cost estimate. In practical application, the final test value of a reliability demonstration test performed during system design would provide the best estimate for MTBF, while solving for the estimated MTTR based on A_C. For this demonstration, the approach selected will suffice. The value for $[C_W]_{Recur}$ was calculated at $151,489.

Table 16

Verification Test Cost for Case Study

Fixed Cost		
	Requirements Definition and Traceability	4000
	Test Plan Development	16000
	Development of Specifications, Manuals, and Procedures	8000
	Test Equipment/Fixture Design	8000
	Test Equipment Fabrication or Procurement	
	Fixed Cost Total:	36000
Recurring Cost		
	Interim Reports	1000
	Final Reports	2500
	Customer Interface	5000
	Technical Community Interface	5100
	Test Readiness Reviews	10000
	Identification of Assets	6000
	Identification of Facilities	1600
	Training	
	Scheduling and/or Procurement of Assets and Facilities	
	Transport/Set-Up/Install/Tear-Down/Return Test Eqp.	7000
	Test Execution	33000
	Simulation Execution	4000
	Data Fusion and Integration	1600
	Post Test Analysis/Reporting	8000
	Engineering Support	8000
	Material Analysis Support	2000
	Acquisition Support	1600
	Post Failure Equipment Repair	12000
	Consumable Items	8000
	Recurring Cost Total:	116400

The case form of the C_{ARM} as specified in equation (48) is

$$C_{ARM} = 36000 + \frac{1}{(1-\alpha')}116400 + 0 + 151489\beta' \quad (54)$$

where α' and β' are defined within time-constrained test plans. The following subsections will demonstrate the time-constrained ADT risk specification process described in Figure 16 for both the FNF and the FTT strategies using the case data.

5.3.1 Case Evaluation for the FNF Strategy

Using the input parameters of U_P = 0.10, D = 2, α = 0.10, and β = 0.10 and equations (22) and (23), the optimal FNF test plan is determined to be (21, 0.1656). Under this test plan, 21 failure repair cycles would be run. At the conclusion, the test metric $z = \sum y_i / \sum x_i$ would be evaluated and the system accepted when z < 0.1656 and rejected otherwise. By equation (34), the expected duration of the test will be 42147 hours. This test plan exceeds the time constraint of 8000 hours specified by the decision maker. Therefore, a time-constrained test plan will be required. For the FNF strategy, equation (36) suggests a *n'* value of 4. Following the process in Figure 17, the next step is to develop alternate test plans under the constraint.

Alternative test plans are generated using equations (38) and (39) with *n'* = 4. Solving for 0 < z < 1 produces the set of alternative test plans shown in Figure 22. Evaluating the C_{ARM} defined by equation (48) produces the cost curves in Figure 23. The low C_{ARM} test plan is defined by the triple (0.224, 0.364, 0.194) which equates to producer and consumer risk statements Producer Risk:

- A system with a true availability of 0.90 will pass the test 77.6% of the time
- Consumer Risk: A system with a true availability of 0.80 will pass the test 36.4% of the time

The decision maker has stated a desire to keep the risk level below 0.30 for both producer and consumer risks. An additional 4000 hours of test time was made available by adding a third test system. A new T'= 12000 is established. An updated *n'* value of 6 is derived through evaluation of equation (36). Calculating the new set of alternative test plans and solving for the associated C_{ARM} results in Figure 24. The low cost alternative is test plan (0.205, 0.292, 0.181). This plan meets the requirement that $\alpha < 0.3$ and $\beta < 0.3$ with a C_{ARM} value of 226698.

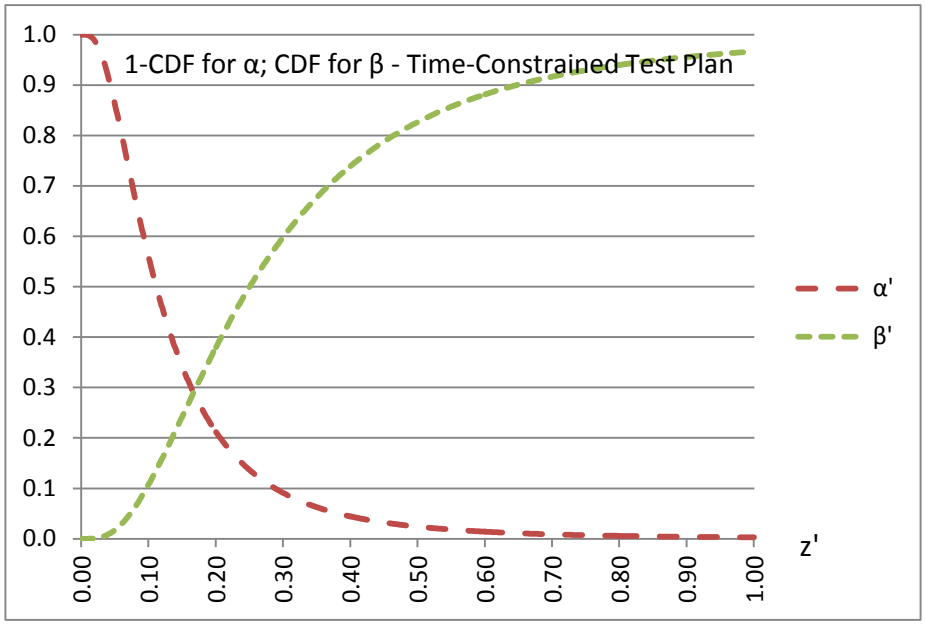

Figure 22. Alternative FNF Test Plans for $A_P = 0.90$, $A_C = 0.80$, and n' = 4 over range $0 < z < 1$

5.3.2 Case Evaluation for the FTT Strategy

Input parameters of $U_P = 0.10$, D = 2, $\alpha = 0.10$, $\beta = 0.10$ and $m_u = 1806$ applied to equations (24) and (25) result in an optimal FTT test plan of (27193, 0.8557). Under this optimal test plan, testing would be run for 27193 hours. At the conclusion of testing, the test metric $A(T) = \sum x_i / (\sum x_i + \sum y_i)$ would be evaluated and compared to the critical

value of 0.8557. The system is accepted if A(T) is greater than 0.8557; otherwise it is rejected.

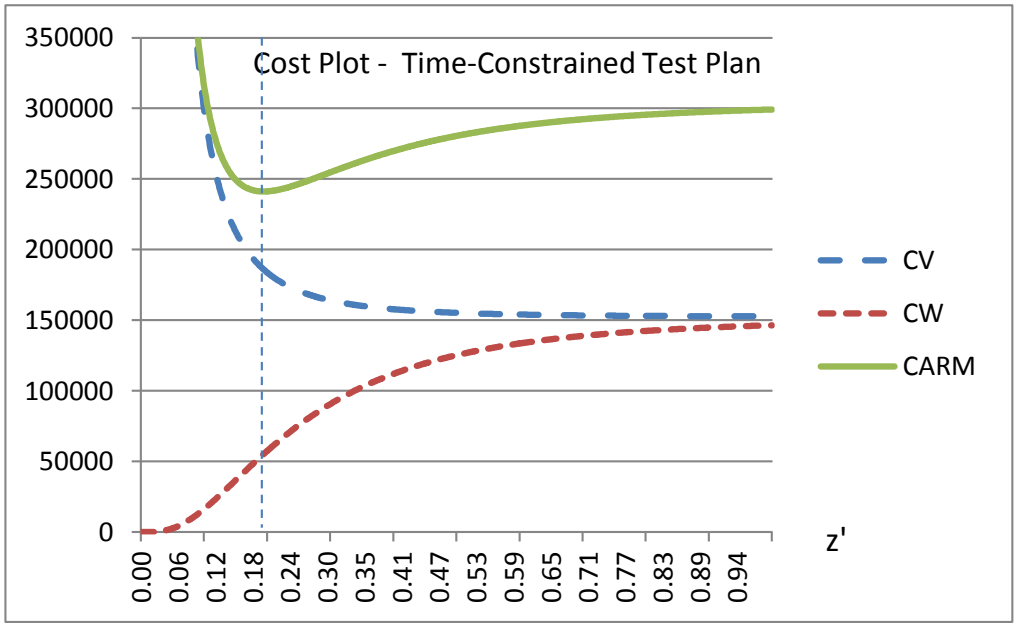

Figure 23. C_{ARM} for alternative FNF Test Plans with $A_P = 0.90$, $A_C = 0.80$, and n' = 4; Cost elements $[C_V]_{Fixed} = 36K$, $[C_V]_{Recur} = 116.4K$, $[C_W]_{Fixed} = 0K$, $[C_W]_{Recur} = 151.5K$

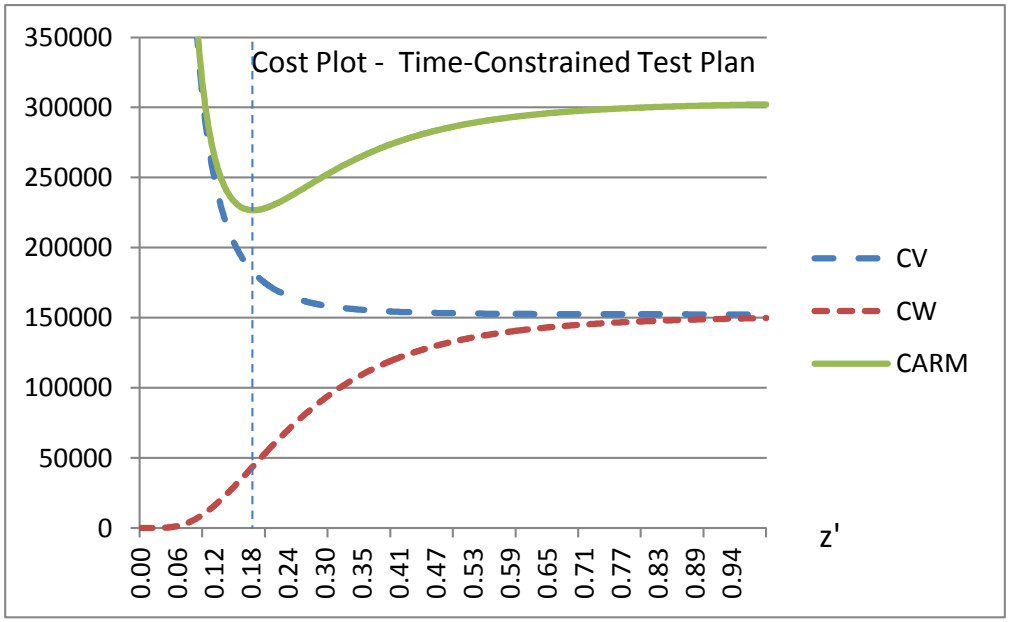

Figure 24. C_{ARM} for alternative FNF Test Plans with $A_P = 0.90$, $A_C = 0.80$, and n' = 6; Cost elements $[C_V]_{Fixed} = 36K$, $[C_V]_{Recur} = 116.4K$, $[C_W]_{Fixed} = 0K$, $[C_W]_{Recur} = 151.5K$

This test plan exceeds the time constraint of 8000 hours specified by the decision maker. Therefore, a time-constrained test plan will be required. The *T'* value for the FTT strategy is set at the constraint value of 8000 hours. Following the process in Figure 17, the next step is to develop alternate test plans under the constraint.

Alternative test plans are generated using equations (40) and (25) with $T' = 8000$. Solving for $0 < \beta < 1$ produces the set of alternative test plans shown in Figure 25. Note the horizontal axis has been reversed from Figure 20 to provide consistency in A(T)' across the test plan and C_{ARM} plots.

Evaluating the C_{ARM} defined by equation (58) produces the cost curves in Figure 26. The low C_{ARM} test plan is defined by the triple (0.170, 0.313, 0.8391) which equates to producer and consumer risk statements

- Producer Risk: A system with a true availability of 0.90 will pass the test 83.0% of the time
- Consumer Risk: A system with a true availability of 0.80 will pass the test 68.7% of the time

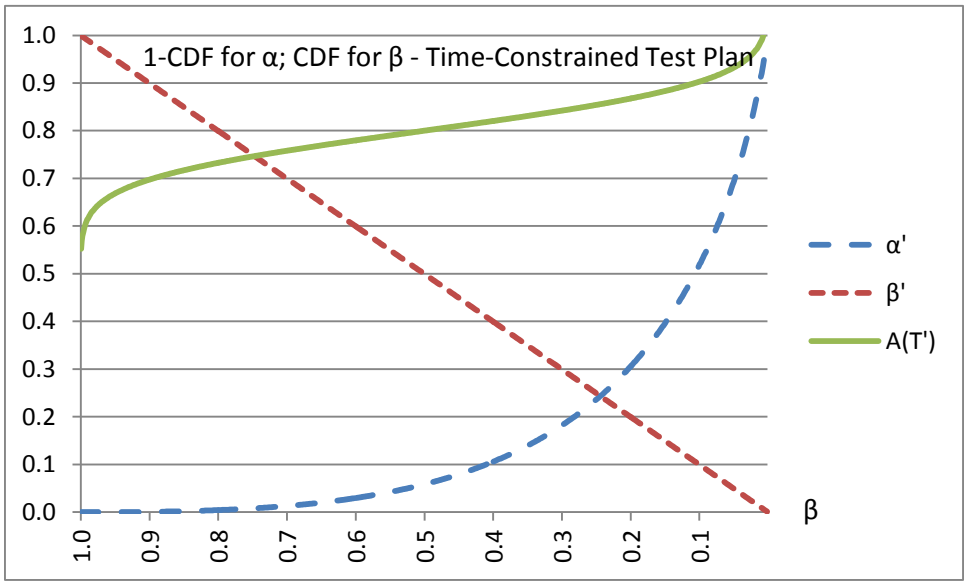

Figure 25. Alternative FTT Test Plans for D = 2, $U_P = 0.10$, $m_u = 1806$, and T' = 8000

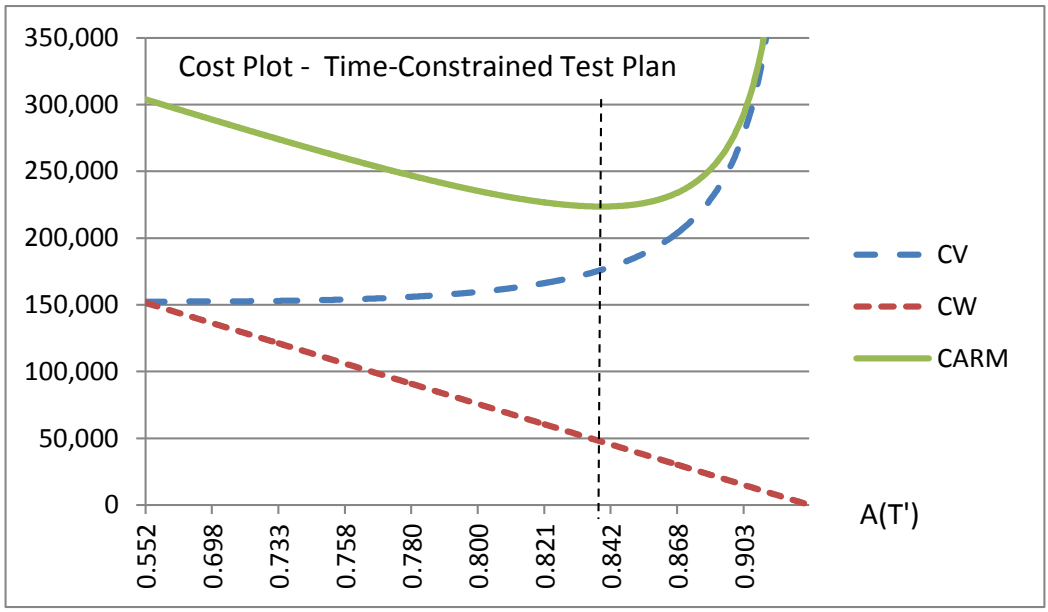

Figure 26. C_{ARM} for FTT Test Plans with D = 2, U_P = 0.10, m_u = 1806, and T' = 8000; Cost elements $[C_V]_{Fixed}$ = 36K, $[C_V]_{Recur}$ = 116.4K, $[C_W]_{Fixed}$ = 0K, $[C_W]_{Recur}$ = 151.5K

The decision maker has stated a desire to keep the risk level below 0.30 for both producer and consumer risks. An additional 4000 hours of test time was made available by adding a third test system. A new T'= 12000 is established. Calculating the new set of alternative test plans and solving for the associated C_{ARM} results in Figure 27.

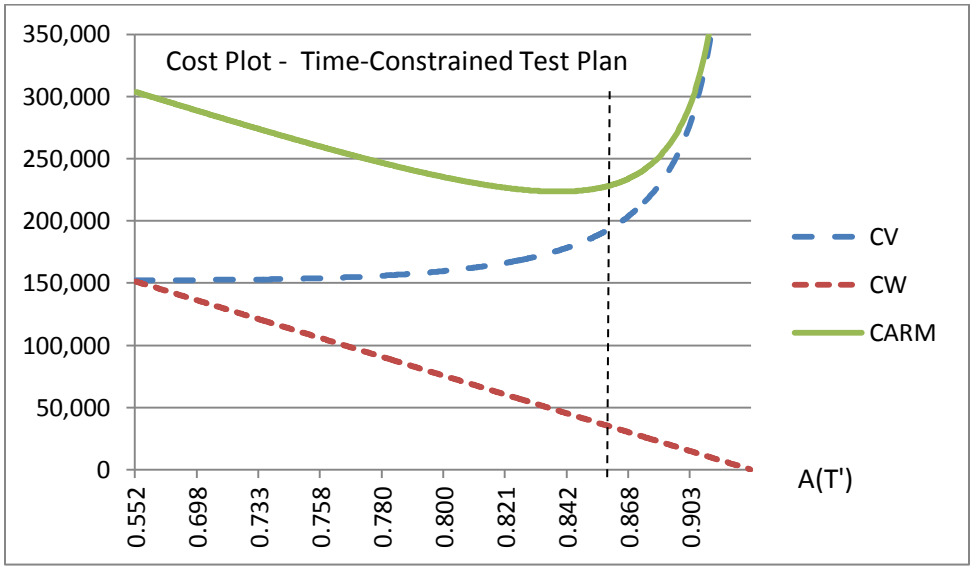

Figure 27. C_{ARM} for FTT Test Plans with D = 2, U_P = 0.10, m_u = 1806, and T' = 12000; Cost elements $[C_V]_{Fixed}$ = 36K, $[C_V]_{Recur}$ = 116.4K, $[C_W]_{Fixed}$ = 0K, $[C_W]_{Recur}$ = 151.5K

The low cost alternative is test plan (0.150, 0.241, 0.8460). This plan meets the requirement that $\alpha < 0.3$ and $\beta < 0.3$ with a C_{ARM} value of 209403.

5.4 Conclusions

The methodology presented in this chapter can be used to manage the risk exposure of a program due to producer and consumer risk growth caused by a time constraint on testing. This method provides a basis for selecting the best risk levels considering the resulting cost of availability risk mitigation. The case study of the verification test design for a surface navy electronics system demonstrated the utility of this approach in providing a decision maker with awareness of changes to risk level caused by time constraint on an optimal test plan. Additionally, the ability to identify a low cost test plan under a specified time constraint was established.

The proposed methodology was effectively demonstrated for both FNF and FTT ADT strategies. Table 17 summarizes the optimal test plans for each strategy. The expected test time is comparable as expected given the time constraint. The C_{ARM} values are within 10%, with the FTT cost the lowest. Interestingly, the FTT risk levels are significantly lower than for FNF. This observation is relevant considering the outcome of Chapter 3. It was determined that FNF had superior quality with respect to actual risk level of a given test plan. However, if the risk levels for FTT are significantly lower than for FNF, then a larger error could still provide a test under the FTT strategy with lower risk levels than for the FNF strategy.

5.5 Future Research

A single area of future research is identified for the application of the methodology presented in this chapter. This area considers the difference in the risk

levels for optimal test plans under the two strategies identified within the case study. Table 17 summarized the results of the case study for both ADT strategies used in time-constrained testing. It was noted that the FNF risk levels are significantly larger than for FTT. In fact, the α' value for FNF is 36.7% larger than that of FTT and the β' value for FNF is 21.2% larger than under the FTT strategy. It is possible these differences are larger than the error differences detected in the Chapter 3 ANOVA.

Table 17

Optimal Test Plans from Time-Constrained ADT Case Study with $A_P = 0.90$, $A_C = 0.80$, $\alpha = 0.10$, and $\beta = 0.10$, $m_u = 1806$, and T' = 12000.

Strategy	α'	β'	Critical Value	C_{ARM}	E[Test Time]
FNF	0.205	0.292	z* = 0.181	226698	12042
FTT	0.150	0.241	A(T)' = 0.8460	209403	12000

Consider the following two conditions

$$(\alpha'_{FNF} - \alpha'_{FTT}) > (\alpha'_{Actual} - \alpha'_{Predicted})_{FTT}$$

$$(\beta'_{FNF} - \beta'_{FTT}) > (\beta'_{Actual} - \beta'_{Predicted})_{FTT}$$

where the left side of the inequalities is the difference in the specific risk levels between the two strategies defined under the same time-constrained test conditions. The right hand side of the inequality is the difference between the actual risk level of the test and that predicted during test design for the FTT strategy. This is the error used to determine the quality measure as defined in previous chapters.

When both equations are true, the FTT strategy provides a more attractive alternative to the decision maker even though the error is larger. This relationship should be evaluated to determine if it tends to be true, or is just the result within a single case study. If the result is common across a wide set of test parameters, FTT could become a more attractive strategy for quality as well as timeliness.

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

The development of complex, large scale systems is required to meet the needs of DoD end users. The maturation and integration of new technology, as well as the implementation of existing technology in new applications, generates technical risk within these development efforts. Demonstration testing is a primary means of mitigating this risk and assuring the consumer that a costly investment will meet stated requirements. The purpose of this research was to investigate and propose a methodology for availability demonstration testing (ADT) that provided users of complex, large scale systems an improved ability to distinguish between high and low availability systems. Further, this method was designed to support a decision maker in the face of a time constraint on test duration.

The consumer's desire to field the end item coupled with the need of the producer to receive timely payment creates pressure to skip the risk reduction step of ADT. This exposes the consumer to the potential of reduced mission success and increased operational support costs. This research defined a methodology for ADT that provides a useful strategy for the consumer to mitigate significant risk without sacrificing the cost of time to field a product or capability.

Development of this method required three significant components which add to the existing body of knowledge. The first was a comparative assessment of performance of ADT strategies to understand if any preferences exist. The next component was the development of an approach for ADT design that was robust against time constraints on the test plan. The third component was the definition of a procedure for ADT design which implements the results of the research into a methodology that provides awareness of risk levels in time-constrained ADT, and offers an analysis of alternatives to select the best sub-optimal test plan.

A design of experiments was conducted using simulation to assess the performance of the three ADT strategies with respect to quality and timeliness. Arena® simulation software was used to assess the strategies in these response areas. Three significant observations were made:

- The mean deviation between estimated and observed total risk levels for the SEQ test strategy is 4.5 times that of the FNF strategy. Furthermore, the mean deviation for the FTT strategy is between 9.5 to 13.7 times the mean of the FNF strategy. A program that is more sensitive to deviations in α and β risk levels than to the duration of testing should use the Fixed Number of Failure strategy for ADT.

- The average test duration for SEQ testing is nearly 1.7 times that of the FTT strategy. The FNF strategy takes 2.5 times longer than FTT. A program that is more sensitive to test duration than to deviation in risk levels should use the Fixed Test Time strategy for ADT optimization.

- The quality response is very sensitive to the discrimination ratio factor. Regardless of the test strategy, the larger the separation between A_P and A_C, the greater the expected error between the actual risk level and the estimated risk level. This sensitivity is greater for FTT than for either FNF or SEQ strategies. Therefore, if the FTT strategy is considered, the test designer must evaluate the discrimination ratio for values in excess of 2.5-3.0, where the quality performance is seen to degrade by over 250%.

Analysis revealed that substitution of a specified time constraint tailored to a format compatible with the FNF or FTT test strategies produced a pair of dependent relationships between risk levels and the critical test value. These relationships could be analyzed in a cost context to compare alternative test plans and choose a low cost alternative. Specific observations were:

- A reduction in test duration while holding availability targets constant has a direct negative effect on both the producer and consumer risk levels for a given ADT plan.

- Application of time constraint to the FNF and FTT strategies created dependent relationships between both risk levels and a common critical test value. A set of alternative test plans were characterized by the triples (α', β', z') for FNF and (α', β', A(T)') for FTT. These sets of alternative test plans were unique for a specified time constraint.

- The cost for availability risk mitigation (C_{ARM}) was defined in terms of type I and type II errors for hypothesis testing under ADT. This integrated cost equation provides a basis to compare alternative test plans under time-constrained FNF and FTT strategies to determine the least cost alternative.

- This method creates a robust basis for program manager to understand changes in risk levels and a good alternative test plan in the event of time constraint to the test and evaluation program.

The research created recognition of the need for a decision support tool to enable a decision maker to understand the changes to α and β as test duration was modified, and make decisions based on the true risk exposure. This understanding led to the development of the time-constrained availability demonstration test risk specification process. A surface navy electronic system case study was used to validate the effectiveness of this process in enabling a decision maker to understand risk levels under time constraint, and to effectively develop an alternative test plan.

6.1 Future Work

Across the three components of this research, several opportunities have been identified for future research. These are summarized below:

- Sequential demonstration test techniques provide potential to reduce test duration. The literature provided no discussion on the evaluation of the probability that test duration under the sequential strategy will exceed the test duration under the alternative strategies. Further expansion on the Sequential test strategy could be accomplished by looking at truncation methods which limit the test time. These approaches retain the advantage of a shorter average test length, while limiting the disadvantage of a significantly larger maximum test time. The use of design of experiments to manage a simulation analysis aimed at evaluating the effect of truncation on producer and consumer risk levels would be valuable.

- The simulation study summarized within this research focused on an exponential failure distribution and an exponential repair distribution. The strategies which define the Method factor levels can be defined in terms of an exponential failure distribution and a gamma repair distribution. Given the exponential has limited applicability to practical real world repair processes, expanding the study to the more powerful gamma repair assumption would be useful to expanding decision guidance.

- The C_{ARM} model considers the fixed and variable cost of validation testing. The variable cost measures the additional fiscal cost of testing when a system meeting user requirements was rejected. There is an additional opportunity cost that was not deliberated. If a system meeting specified availability requirements is rejected and subject to retest, a time delay in fielding the system is created. There is an opportunity cost associated with this time delay that should be incorporated within the cost model.

- The closing paragraphs of Chapter 5 discuss the comparison of results from the case study for FNF verses FTT strategies. Within that discussion, the point is raised that the more attractive risk levels of the FTT strategy as compared to those obtained with the FNF strategy would offset the larger error seen in the ANOVA assessment in Chapter 3. It would be worthwhile to consider a more direct evaluation of these strategies under time constraint to determine if the same conclusions would be reached concerning preference of strategy with respect to quality.

REFERENCES

British Standards Institution. (1993). *Reliability of Systems, Equipment, and Components – Part 10: Guide to Reliability Testing, Section 10.3 Compliance test procedures for steady-state availability* (BS 5760-10.3:1993), London, UK.

Chairman of the Joint Chiefs of Staff (CJCS), (2012). *Joint Capabilities Integration and Development Systems* (CJCSI 3170.01H). Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=267681

Department of Defense. (1982). *Test and Evaluation of System Reliability, Availability, and Maintainability – A Primer* (DoD 3235.1-H). Retrieved from Barringer & Associates, Inc website: http://www.barringer1.com/mil_files/DOD3235.1-H.pdf

Department of Defense. (2005). *DoD Guide for Achieving Reliability, Availability, and Maintainability*. Retrieved from Defense Acquisition University website: https://acc.dau.mil/CommunityBrowser.aspx?id=378067

Department of Defense. (2008). *Defense and Program-unique Specifications Format and Content* (MIL-STD-691E w/change 1). Retrieved from Defense Acquisition University website: https://acc.dau.mil/CommunityBrowser.aspx?id=378460

Director, Operational Test and Evaluation. (2012). *2012 Annual Report to Congress*. Retrieved from Office of Director, Operational Test and Evaluation website: http://www.dote.osd.mil/pub/reports/FY2012/

Ebeling, C.E. (2010). *An Introduction to Reliability and Maintainability Engineering, Second Edition*. Long Grove IL: Waveland Press.

Federal Aviation Administration. (2008). *Reliability, Maintainability, and Availability (RMA) Handbook* (FAA-HDBK-006A). Retrieved from Quanterion Solutions Incorporated website: http://quanterion.com/Publications/FAA-HDBK-006A/index.asp

Fu, K., Yu, Y., Zhang, L., and Xu, Y. (2012). *Study on Data Usability Judgment Method on Inherent Availability Demonstration Test*, Proceedings of the 2012 Prognostics & Health Management Conference, Beijing, China.

General Accountability Office. (2011) *Report to Congressional Committees: Defense Acquisitions – Assessments of Selected Weapon Programs*. Retrieved from Government Accountability Office website: http://www.gao.gov/new.items/d11233sp.pdf

General Accountability Office. (2013) *Report to Congressional Committees: Defense Acquisitions – Assessments of Selected Weapon Programs*. Retrieved from Government Accountability Office website: http://www.gao.gov/assets/660/653379.pdf

Gray, H. L., and Schucany, W. R. (1969). Lower Confidence Limits for Availability Assuming Lognormally Distributed Repair Times. *IEEE Transactions on Reliability*, R-18, No 4, 157-162.

Hallgren, B. (1986). Availability Compliance Testing of Systems with Long Mean Time Between Failures. *Reliability Engineering*, 15, 83-94.

Institute of Electrical and Electronic Engineers. (2011). *Systems Engineering-Application and Management of the Systems Engineering Process.* (IEEE Std-1220-2005), New York, NY.

Joint Requirements Oversight Council (JROC). (2012). *Manual for the Operation of the Joint Capabilities Integration and Development Systems.* Retrieved from Defense Acquisition University website: https://dap.dau.mil/policy/Documents/2012/JCIDS%20Manual%2019%20Jan%202012.pdf

Kelton, W.D., Sadowski, R.P., and Swets, N. B (2010). *Simulation with Arena, Fifth Edition*. New York NY: McGraw Hill.

Kleyner, A., Sandborn, P., and Boyle, J. (2004). Minimization of Life Cycle Cost Through Optimization of the Validation Program – A Test Sample Size and Warranty Cost Approach. *Proceedings of the 2004 Annual Reliability and Maintainability Symposium*, 553-558.

Kleyner, A., and Sandborn, P. (2008). Minimizing Life Cycle Cost by Managing Product Reliability via Validation Plan and Warranty Return Cost. *International Journal Production Economics*, 12, 796-807.

Leung, H. K. N., and White, L. (1991). A Cost Model to Compare Regression Test Strategies. *Proceedings of the 1991 IEEE Conference on Software Maintenance*, 201-208.

National Aeronautical and Space Administration. (1998). *Planning, Developing and Managing an Effective Reliability and Maintainability (R&M) Program* (NASA-STD-8729.1). Retrieved from National Aeronautical and Space Administration website: http://www.hq.nasa.gov/office/codeq/doctree/87291.htm

Pell, C. A., Hall R. C., and Schneider, R. C. (1978). Test Methods for Steady State Availability. *Proceedings of the 1978 Annual Reliability and Maintainability Symposium*, 167-172.

Rise, J. (1979), Compliance Test Plans for Availability. *Proceedings of the 1979 Annual Reliability and Maintainability Symposium*, 368-373.

Schafer, R. E. and Takenaga, R. (1972). A Sequential Probability Ratio Test for Availability. *Technometrics*, 14-1, 123-135.

Sgarbossa, F., and Pham, H. (2010). A Cost Analysis of Systems Subject to Random Field Environments and Reliability. *IEEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Reviews*, 40-4, 429-437.

Thompson, M. (1966). Lower Confidence Limits and a Test of Hypotheses for System Availability. *IEEE Transactions on Reliabililty*, R-15, No. 1, 32-36.

UK Ministry of Defence, DES JSC TLS POL REL. (2011). *Applied R&M Manual, for Defence Systems (GR-77 Issue 2009), Part C, Chapter 40, Reliability Demonstration*. MoD Abbey Wood South, Bristol.

U. S. Army Electronics Command. (1975). *Life Cycle Cost Model (Research and Development Technical Report ECOM-4338*. Fort Monmouth, NJ.

Usher, J.S., and Taylor, G.D. (2006). Availability Demonstration Testing. *Quality and Reliability Engineering International*, 22, 473-479.

APPENDIX A: LIST OF ABBREVIATIONS

ADT	Availability Demonstration Testing
DoD	Department of Defense
FAA	Federal Aviation Administration
FNF	Fixed Number of Failures
FoS	Family of Systems
FTT	Fixed Test Time
JCIDS	Joint Capabilities Integration and Development System
KPP	Key Performance Parameter
KSA	Key System Attribute
MDT	Mean Down Time
MTBF	Mean Time Between Failure
MTBM	Mean Time Between Maintenance
MTR	Mean Time to Restore
MTTR	Mean Time to Repair
NASA	National Aeronautical and Space Administration
OC	Operating Characteristic
RMA	Reliability, Maintainability, and Availability
SEQ	Sequential Testing
SoS	System of Systems
TAFT	Test, Analyze, Fix and Test

APPENDIX B: NOTATION

α	the nominal producer's risk (type I risk)
β	the nominal consumer's risk (type II risk)
$L_A(i)$	the acceptance limit for stage i in a sequential demonstration test
$L_R(i)$	the rejection limit for stage i in a sequential demonstration test
D	the discrimination ratio, $D = U_P/U_C$
U_P	the specified acceptable value of steady-state unavailability, $U_P=(1-A_P)$
U_C	the lowest acceptable/highest unacceptable value of steady-state unavailability, $U_C=(1 - A_C)$
Z_{crit}	the rejection limit in a compliance test
$z_{1-\alpha}$	the 1-α fractile of the standardized normal distribution
A_P	the specified acceptable value of steady-state availability
A_C	the lowest acceptable/highest unacceptable value of steady-state availability
A(T)*	critical test value for a fixed test time test
$F_{1-\alpha}(\nu1, \nu2)$	the 1-α fractile of the F-distribution with degrees of freedom $\nu1$ and $\nu2$.
i	index variable for stage number in sequential testing
m_d	the mean down time
m_u	the mean up time

n	the sample size, i.e. the number of failures, for a fixed number of failures test; the number of failure/repair cycle in a sequential test
p	shape parameter for the gamma distribution (used for time to repair distribution)
T*	the duration for a fixed time test
X	the accumulated up time, X=∑xi
x_i	up time number i, i=1,2,3,…
Y	the accumulated up time, Y=∑yi
y_i	down time number i, i=1,2,3,…
Z	the ratio of downtime to uptime, Z=Y/X

For distributions (from Kelton, Sadowski, and Swets, 2010):

Gamma(p,θ) - $f(t) = \frac{\theta^{-p} t^{p-1} e^{-\frac{t}{\theta}}}{\Gamma(p)}$ for t > 0

where $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$ and Γ(p) = (p-1)! for p an integer

Mean = pθ; Variance = $p\theta^2$

Exponential(θ) - $f(t) = \left[\frac{1}{\theta} e^{-\frac{t}{\theta}}\right]$ for t > 0

Mean = θ; Variance = θ^2

Note, Gamma(1,θ) is equivalent to Exponential(θ).

APPENDIX C: CHAPTER 3 SIMULATION, DESIGN OF EXPERIMENTS AND ANOVA ARTIFACTS

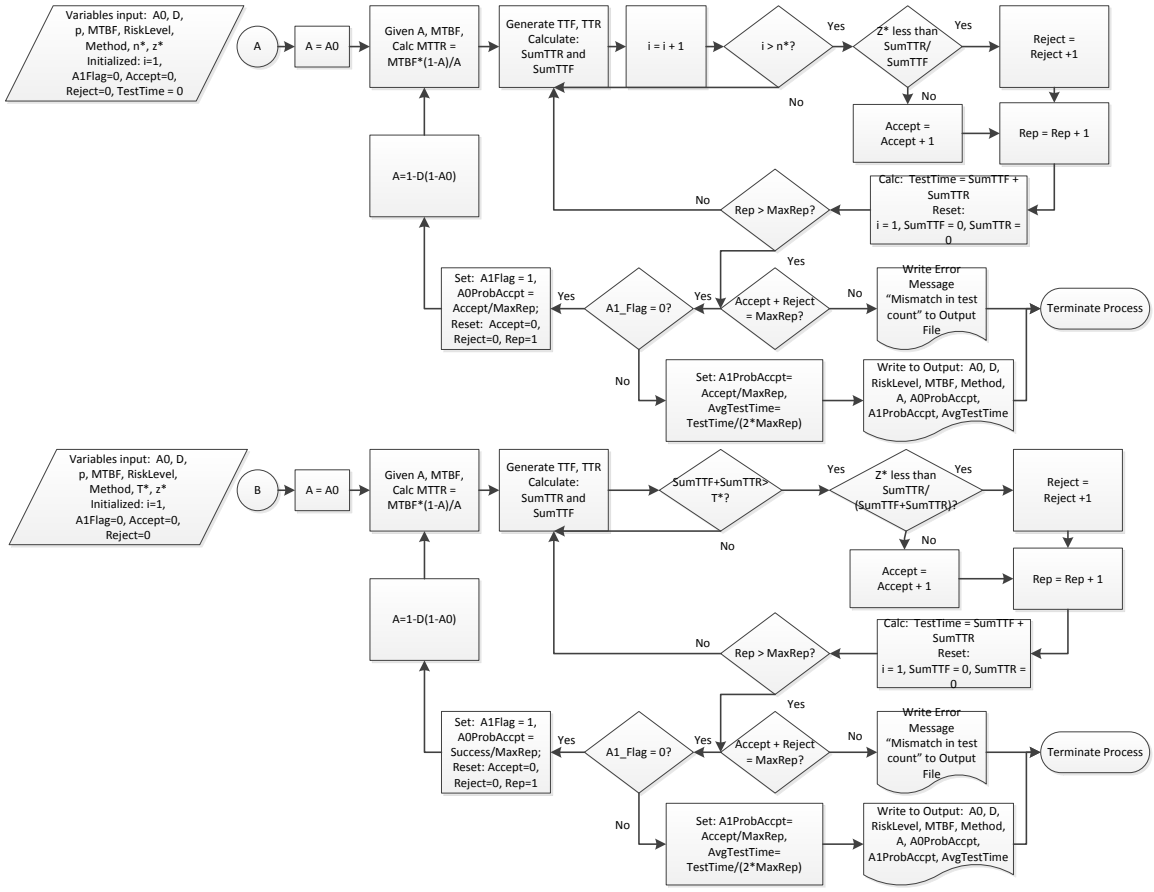

Figure C-1. Flow diagram for Fixed Number of Failure and Fixed Test Time modules

Table C-1

Experimental Runs for ADT Comparison Design of Experiments (First 36 of 324 runs)

Exp_No	DOE_Rep	A0	D	MTBF	RiskLevel	Method	Zcrit	Nstar	Tstar	p_Value
1	2	0.9	2	2000	0.01	4				1
2	2	0.8	4	1	0.01	3	0.6		2.165	1
3	2	0.9	4	1	0.01	2	0.2286		5.893	1
4	1	0.95	4	2000	0.1	4				1
5	1	0.95	2	2000	0.01	1	0.0763	79		1
6	2	0.95	2	1	0.1	4				1
7	2	0.95	4	100	0.05	1	0.1118	10		1
8	1	0.95	4	1	0.01	3	0.1059		8.232	1
9	1	0.9	2	100	0.01	2	0.1443		4962.7	1
10	1	0.95	4	1	0.05	1	0.1118	10		1
11	2	0.8	4	2000	0.1	3	0.6		1313.9	1
12	1	0.95	4	2000	0.01	3	0.1059		16464	1
13	2	0.8	4	1	0.1	1	0.763638	3		1
14	2	0.9	2	100	0.1	3	0.1443		1506	1
15	1	0.8	2	2000	0.05	1	0.4081	23		1
16	1	0.8	4	1	0.05	4				1
17	2	0.8	4	1	0.05	2	0.6		1.082	1
18	2	0.8	4	2000	0.01	2	0.6		4330	1
19	1	0.95	4	2000	0.05	2	0.1059		8230	1
20	1	0.8	4	100	0.01	2	0.6		216.5	1
21	1	0.8	2	100	0.05	2	0.2971		1838.2	1
22	2	0.9	2	1	0.01	1	0.1664	67		1
23	2	0.8	2	2000	0.1	4				1
24	2	0.8	4	1	0.1	4				1
25	2	0.8	2	2000	0.1	3	0.2971		22316	1
26	2	0.9	2	100	0.1	4				1
27	2	0.9	4	100	0.05	4				1
28	1	0.8	2	1	0.1	2	0.2971		11.158	1
29	1	0.8	4	2000	0.1	1	0.763638	3		1
30	2	0.95	4	1	0.05	4				1
31	1	0.95	2	2000	0.01	2	0.0714		112586	1
32	2	0.95	2	1	0.01	1	0.0763	79		1
33	1	0.9	2	100	0.01	4				1
34	2	0.95	2	1	0.01	3	0.0714		56.293	1
35	1	0.95	4	1	0.05	3	0.1059		4.115	1
36	1	0.8	2	1	0.01	3	0.2971		36.769	1

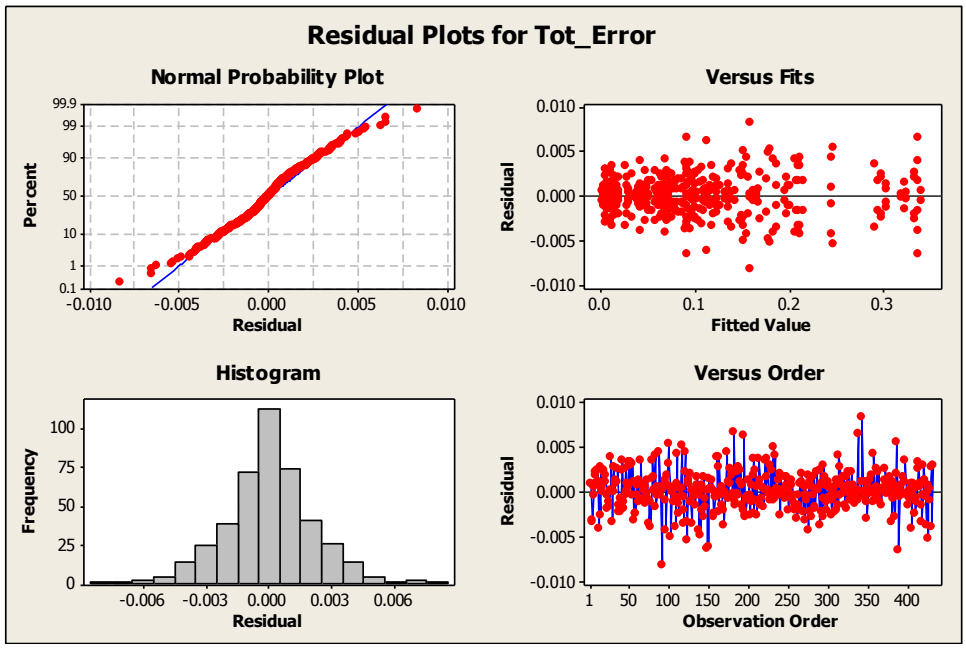

Figure C-2. Residual Analysis for Quality Response:
Model is A0| D| MTBF| RiskLevel| Method

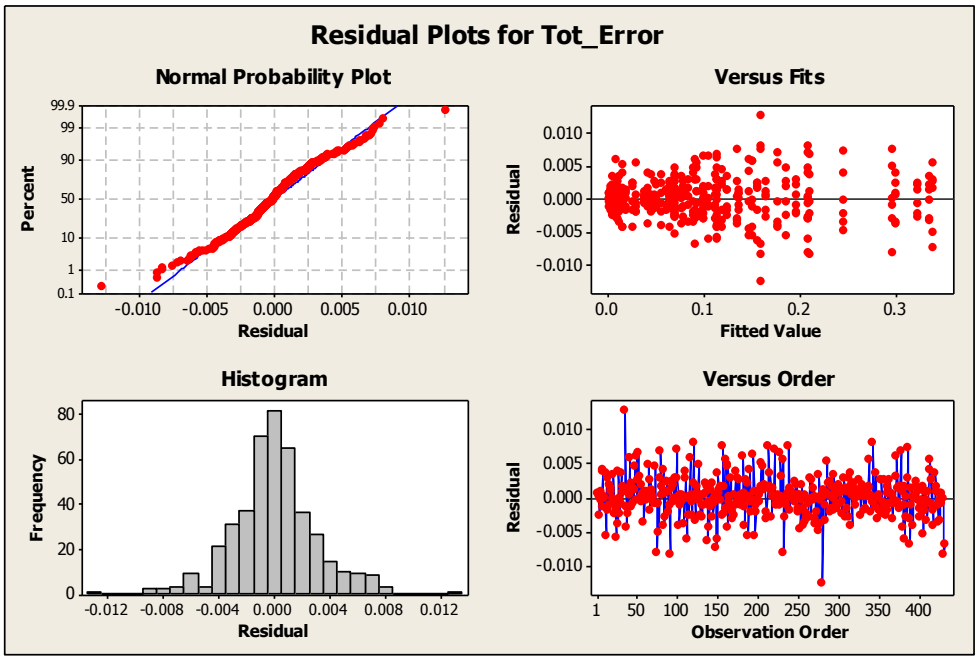

Figure C-3. Residual Analysis for Quality Response:
Model is A0| D| RiskLevel| Method

Table C-2

General Linear Model: Tot_Error versus A0, D, MTBF, RiskLevel, Method (Quality Response)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
D	1	0.580734	0.580734	0.580734	64458.19	0.000
Method	3	1.710063	1.710063	0.570021	63269.11	0.000
D*Method	3	0.475943	0.475943	0.158648	17608.99	0.000
RiskLevel	2	0.198814	0.198814	0.099407	11033.63	0.000
A0	2	0.068842	0.068842	0.034421	3820.56	0.000
RiskLevel*Method	6	0.068475	0.068475	0.011413	1266.72	0.000
A0*D*RiskLevel	4	0.025076	0.025076	0.006269	695.82	0.000
A0*RiskLevel*Method	12	0.067655	0.067655	0.005638	625.78	0.000
A0*D*RiskLevel*Method	12	0.064985	0.064985	0.005415	601.08	0.000
A0*Method	6	0.020521	0.020521	0.00342	379.61	0.000
A0*D	2	0.006528	0.006528	0.003264	362.31	0.000
A0*RiskLevel	4	0.010537	0.010537	0.002634	292.38	0.000
D*RiskLevel*Method	6	0.014741	0.014741	0.002457	272.70	0.000
A0*D*Method	6	0.012657	0.012657	0.002109	234.13	0.000
D*RiskLevel	2	0.001138	0.001138	0.000569	63.15	0.000
A0*D*MTBF*RiskLevel*Method	24	0.000391	0.000391	0.000016	1.81	0.015
A0*D*MTBF*RiskLevel	8	0.000172	0.000172	0.000021	2.38	0.018
A0*MTBF*RiskLevel*Method	24	0.000346	0.000346	0.000014	1.60	0.043
MTBF*RiskLevel	4	0.000082	0.000082	0.00002	2.27	0.063
D*MTBF*RiskLevel*Method	12	0.000186	0.000186	0.000016	1.72	0.063
MTBF	2	0.000044	0.000044	0.000022	2.42	0.091
D*MTBF*RiskLevel	4	0.00007	0.00007	0.000018	1.95	0.103
A0*MTBF	4	0.000044	0.000044	0.000011	1.21	0.308
A0*D*MTBF*Method	12	0.000121	0.000121	0.00001	1.12	0.347
D*MTBF*Method	6	0.000052	0.000052	0.000009	0.97	0.447
A0*MTBF*RiskLevel	8	0.000068	0.000068	0.000009	0.95	0.479
MTBF*Method	6	0.000046	0.000046	0.000008	0.86	0.528
A0*D*MTBF	4	0.000024	0.000024	0.000006	0.67	0.610
D*MTBF	2	0.000007	0.000007	0.000004	0.41	0.667
MTBF*RiskLevel*Method	12	0.000075	0.000075	0.000006	0.69	0.756
A0*MTBF*Method	12	0.00006	0.00006	0.000005	0.55	0.877
Error	216	0.001946	0.001946	0.000009		
Total	431	3.330442				

Table C-3

General Linear Model: Tot_Error versus A0, D, RiskLevel, Method (Quality Response)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
D	1	0.580734	0.580734	0.580734	55986.48	0.000
Method	3	1.710063	1.710063	0.570021	54953.67	0.000
D*Method	3	0.475943	0.475943	0.158648	15294.65	0.000
RiskLevel	2	0.198814	0.198814	0.099407	9583.48	0.000
A0	2	0.068842	0.068842	0.034421	3318.42	0.000
RiskLevel*Method	6	0.068475	0.068475	0.011413	1100.24	0.000
A0*D*RiskLevel	4	0.025076	0.025076	0.006269	604.36	0.000
A0*RiskLevel*Method	12	0.067655	0.067655	0.005638	543.53	0.000
A0*D*RiskLevel*Method	12	0.064985	0.064985	0.005415	522.08	0.000
A0*Method	6	0.020521	0.020521	0.00342	329.72	0.000
A0*D	2	0.006528	0.006528	0.003264	314.69	0.000
A0*RiskLevel	4	0.010537	0.010537	0.002634	253.95	0.000
D*RiskLevel*Method	6	0.014741	0.014741	0.002457	236.86	0.000
A0*D*Method	6	0.012657	0.012657	0.002109	203.36	0.000
D*RiskLevel	2	0.001138	0.001138	0.000569	54.85	0.000
Error	360	0.003734	0.003734	0.00001		
Total	431	3.330442				

Table C-4

Grouping Information for Factor Discrimination Ratio Using Tukey Method and 95.0% Confidence (Quality Response)

D	N	Mean	Grouping
4	216	0.129871	A
2	216	0.056542	B

Table C-5

Grouping Information for Factor Method Using Tukey Method and 95.0% Confidence (Quality Response)

Method	N	Mean	Grouping
2	108	0.178248	A
3	108	0.123700	B
4	108	0.057773	C
1	108	0.013104	D

Table C-6

Grouping Information for Interaction Discrimination Ratio and Method Using Tukey Method and 95.0% Confidence (Quality Response)

Grouping Information Using Tukey Method and 95.0% Confidence

D	Method	N	Mean	Grouping
4	2	54	0.266102	A
4	3	54	0.167974	B
2	2	54	0.090394	C
2	3	54	0.079426	D
4	4	54	0.065109	E
2	4	54	0.050437	F
4	1	54	0.020298	G
2	1	54	0.005909	H

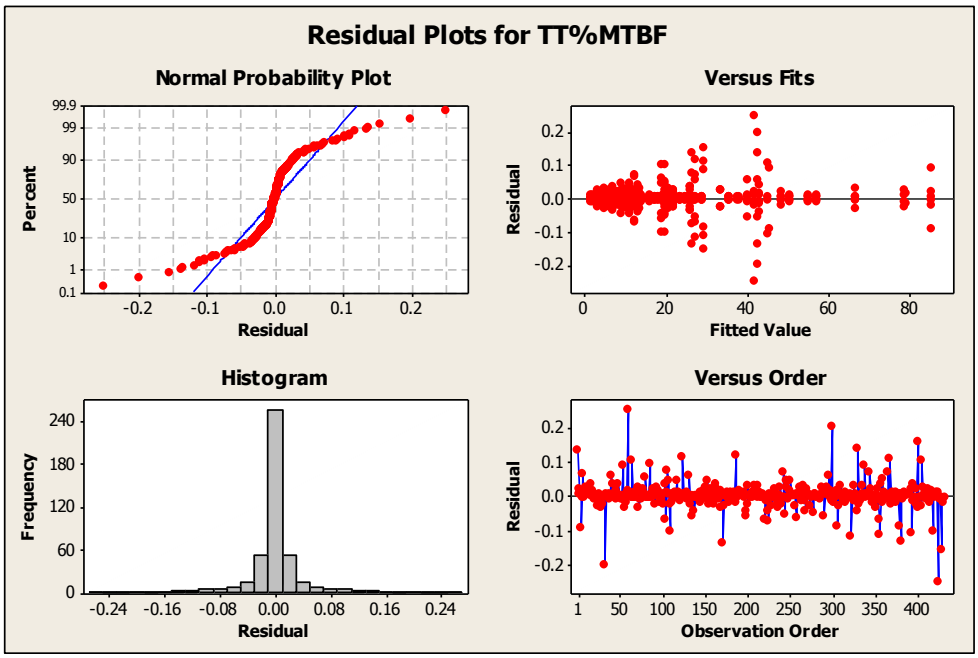

Figure C-4. Residual Analysis for Timeliness Response:
Model is A0| D| MTBF | RiskLevel| Method

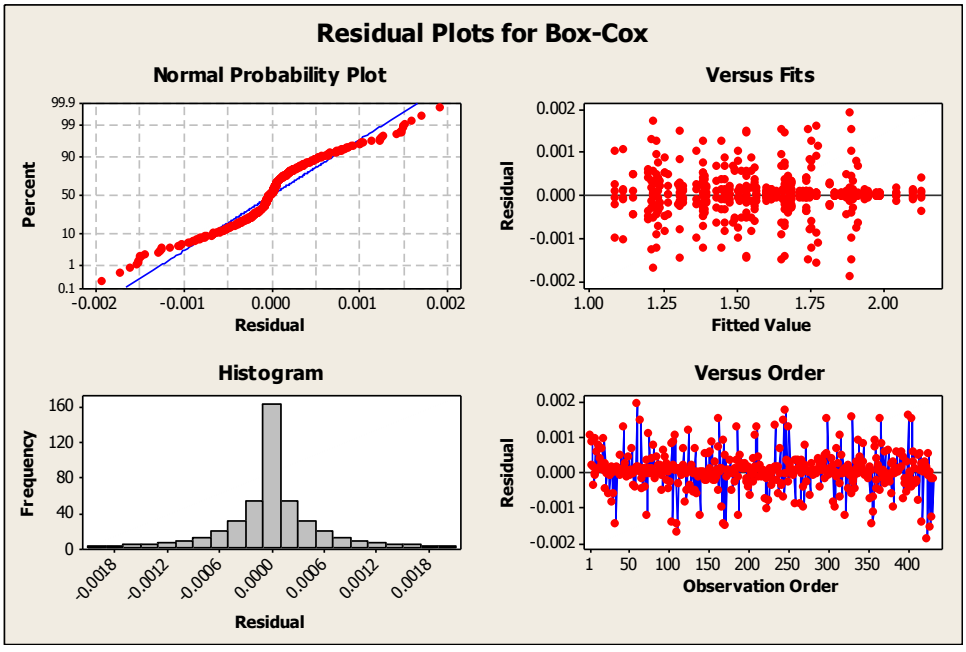

Figure C-5. Residual Analysis for Transformed Timeliness Response:
Model is A0| D| MTBF | RiskLevel| Method

Table C-7

General Linear Model: Box-Cox Transformed Test Time as % of MTBF versus A0, D, MTBF, RiskLevel, Method (Transformed Timeliness Response)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
D	1	17.07952	17.07952	17.07952	29907815.29	0.000
RiskLevel	2	4.98201	4.98201	2.49100	4361977.37	0.000
Method	3	3.98275	3.98275	1.32758	2324723.02	0.000
D*Method	3	0.81714	0.81714	0.27238	476959.76	0.000
D*RiskLevel	2	0.32751	0.32751	0.16376	286751.41	0.000
A0*D	2	0.19668	0.19668	0.09834	172205.64	0.000
A0*Method	6	0.34885	0.34885	0.05814	101810.03	0.000
A0	2	0.07105	0.07105	0.03553	62210.53	0.000
RiskLevel*Method	6	0.14299	0.14299	0.02383	41731.30	0.000
A0*RiskLevel	4	0.06901	0.06901	0.01725	30211.99	0.000
D*RiskLevel*Method	6	0.05272	0.05272	0.00879	15387.62	0.000
A0*D*RiskLevel	4	0.03379	0.03379	0.00845	14790.76	0.000
A0*D*Method	6	0.01984	0.01984	0.00331	5790.73	0.000
A0*RiskLevel*Method	12	0.03846	0.03846	0.00320	5612.05	0.000
A0*D*RiskLevel*Method	12	0.03320	0.03320	0.00277	4845.39	0.000
A0*D*MTBF*Method	12	0.00001	0.00001	0.00000	2.05	0.021
A0*MTBF	4	0.00001	0.00001	0.00000	2.30	0.060
MTBF*RiskLevel	4	0.00000	0.00000	0.00000	1.63	0.167
A0*MTBF*RiskLevel*Method	24	0.00002	0.00002	0.00000	1.21	0.232
MTBF	2	0.00000	0.00000	0.00000	1.18	0.309
A0*MTBF*Method	12	0.00001	0.00001	0.00000	1.10	0.360
A0*MTBF*RiskLevel	8	0.00000	0.00000	0.00000	1.07	0.388
MTBF*RiskLevel*Method	12	0.00001	0.00001	0.00000	1.03	0.423
D*MTBF	2	0.00000	0.00000	0.00000	0.62	0.538
A0*D*MTBF*RiskLevel	8	0.00000	0.00000	0.00000	0.85	0.557
D*MTBF*Method	6	0.00000	0.00000	0.00000	0.80	0.572
D*MTBF*RiskLevel	4	0.00000	0.00000	0.00000	0.64	0.638
D*MTBF*RiskLevel*Method	12	0.00001	0.00001	0.00000	0.78	0.675
A0*D*MTBF	4	0.00000	0.00000	0.00000	0.53	0.715
A0*D*MTBF*RiskLevel*Method	24	0.00001	0.00001	0.00000	0.78	0.756
MTBF*Method	6	0.00000	0.00000	0.00000	0.28	0.945
Error	216	0.00012	0.00012	0.00000		
Total	431	28.19574				

Table C-8

Grouping Information for Factor Discrimination Ratio Using Tukey Method and 95.0% Confidence (Transformed Timeliness Response)

D	N	Mean	Grouping
2	216	1.784	A
4	216	1.386	B

Table C-9

Grouping Information for Factor Risk Level Using Tukey Method and 95.0% Confidence (Transformed Timeliness Response)

RiskLevel	N	Mean	Grouping
0.01	144	1.725	A
0.05	144	1.565	B
0.10	144	1.465	C

Table C-10

Grouping Information for Factor Method Using Tukey Method and 95.0% Confidence (Transformed Timeliness Response)

Method	N	Mean	Grouping
1	108	1.715	A
4	108	1.634	B
3	108	1.523	C
2	108	1.468	D

Table C-11

Method	Mean Response	Untransformed Response (x^5.8824)	Ratio to Best
$FTT^{(-)}$	1.468	9.566	1:1
$FTT^{(+)}$	1.523	11.877	1.24:1
SEQ	1.634	17.965	1.88:1
FNF	1.715	23.880	2.50:1

CURRICULUM VITAE

RANDALL L. WALKER
4120 South Purple Finch Drive
Bloomington, Indiana 47403
Tel: (812) 327-2977
Email: randall.walker@navy.mil

EDUCATION:

University of Louisville, Louisville, KY
PhD Candidate in Industrial Engineering, Exp. Graduation: Dec 2013
Specialization: Reliability, Maintainability, and Availability
Advisor: Dr. John S. Usher
Dissertation: 'A Cost Model for Managing Producer and Consumer Risk in Availability Demonstration Testing'
M.Eng. in Industrial Engineering, Aug 1991
Specialization: Operations Research
Advisor: Dr. William E. Biles
Thesis: 'Minimizing Time to Market in Plastic Product Development Utilizing Concurrent Engineering'
B.S. in Engineering Science, Aug 1990
B.S. in Mathematics, Dec 1988

RESEARCH INTERESTS

1. Life Cycle Cost Modeling for State of the Art Technology
2. Reliability, Maintainability, and Availability Modeling and Optimization

Current Research: Optimization of validation test planning under time constraints

JOURNAL ARTICLES

DePuy, G.W., Usher, J.S., Walker, R. L., and Taylor, G. D., "Production Planning for Remanufactured Products", *Production Planning and Control*, Vol 18, No. 7, pp 573-583.
*Note: This paper received the *Production Planning and Control* Best Paper Award for 2007.

CONFERENCE

DePuy, G.W., Usher, J.S., Arterburn, B., Walker, R., and Fredrick, M.,

PROCEEDINGS

"Workforce training schedule for logistics skills", The Proceedings of the 2006 Industrial Engineering Research Conference, Orlando, FL, May 20-24, 1999, CD-ROM 6 pages.

Walker, Randall L., and Usher, J. S., "Operational Availability of Complex Navy Multi-function Systems", The Proceedings of the Eighth Industrial Engineering Research Conference, Phoenix, AZ, May 15-16 1999, CD-ROM.

Walker, Randall L., and Brown, T., "Reliability Improvement of Components through Warranty and Incentives", The Proceedings of the Eighth Industrial Engineering Research Conference, Phoenix, AZ, May 15-16 1999, CD-ROM.

Usher, John S., and Walker, R., "Incomplete Field Data: The Effect on Component Reliability Estimation", The Proceedings of The Sixth Industrial Engineering Research Conference, Miami Beach, FL, May 17-18, 1997, pp 24-29.

Walker, Randall and Sims, S., "Acquisition Engineering for AWS Microwave Tubes - Tube Improvement Program", Proceedings of the Fleet Maintenance Symposium, San Diego, CA, 19-21 October 1993, pp 127-143.

Biles, W. E., Walker, R. L. and Gornet, T. J, "Concurrent Engineering of New Plastic Products", Proceedings of the International Conference on Computer Integrated Manufacturing - ICCIM91, Singapore, October 1991, pp. 191-194.

ACADEMIC EXPERIENCE

Instructor, Mathematics Department, Indiana Vocational Technical College, Bloomington, IN

Intermediate Algebra: Prepared material, lectured, and evaluated students based on competency displayed through tests, quizzes, and homework assignments.

WORK EXPERIENCE

Naval Surface Warfare Center, Crane Division, Crane, IN.
Senior Industrial Engineer

PhD Fellow: University of Louisville, 2011-2013
Division Manager: Expeditionary Electronics Warfare Systems Division, 2008-2011
Reengineering Lead: Microwave Systems Department, 2006-2008
Branch Manager: Solid State Devices Branch, 2001-2006
Deputy Readiness and Logistics Support Manager: DD-21 Shipbuild Program, 1997-2001
Project Manager: Microwave Technology Project, 1996-1997.
RMA Engineer: 1991-1996.