INSTITUTE FOR DEFENSE ANALYSES

DOT&E Reliability Course

Matthew Avery
Jonathan Bell
Rebecca Dickinson
Laura Freeman

June 2016

Approved for public release.

IDA Document NS D-5836
Log: H 16-000799

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

IDA

The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

About This Publication

This reliability course provides information to assist DOT&E action officers in their review and assessment of system reliability. Course briefings cover reliability planning and analysis activities that span the acquisition life cycle. Each briefing discusses review criteria relevant to DOT&E action officers based on DoD policies and lessons learned from previous oversight efforts.

Copyright Notice

© 2016 Institute for Defense Analyses

4850 Mark Center Drive, Alexandria, Virginia 22311-1882 • (703) 845-2000.

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (a)(16) [Jun 2013].

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-5836

DOT&E Reliability Course

Matthew Avery
Jonathan Bell
Rebecca Dickinson
Laura Freeman

Executive Summary

This reliability course provides information to assist DOT&E action officers in their review and assessment of system reliability, focusing on review criteria relevant to the DOT&E action officers based on DoD policies and lessons learned from previous oversight efforts. The course consists of seven briefings that cover reliability planning and analysis activities that span the acquisition life cycle, including:

- An introduction that provides an overview of the course and motivation for improving system reliability.
- A Reliability, Availability, Maintainability (RAM) Requirements Review briefing that highlights the importance of reviewing RAM requirements early in the program's lifecycle and the criteria that should be considered during the review process.
- An overview of the importance and process of reliability growth planning, focusing on information essential to support review of Test and Evaluation Master Plans (TEMP) and test plans. The briefing also describes how to use typical reliability planning growth models.
- A briefing on the importance of design reviews in the Reliability Growth Planning process that focuses on the relevant questions to consider during design reviews.
- Detail on how programs should document their reliability growth plan in the TEMP including a discussion on criteria that should be considered during the review process and how to assess the adequacy of an OT to evaluate reliability.
- A briefing that focuses on analysis of reliability in Developmental Testing (DT) that provides an overview of DT activities that are essential to support reliability assessment and tracking and explains how to determine if the proposed DT will be adequate.
- Detail on how to analyze RAM data for DOT&E reports using common methods such as development of confidence bounds, analysis of censored data, comparison to legacy systems, estimation of the reliability growth potential, and subsystem failure analysis.

DOT&E Reliability Course

Catherine Warner
Matthew Avery
Jonathan Bell
Rebecca Dickinson
Laura Freeman

2 June 2016

Course Objective and Overview

Objective

- Provide information to assist DOT&E action officers in their review and assessment of system reliability.

Overview and Agenda

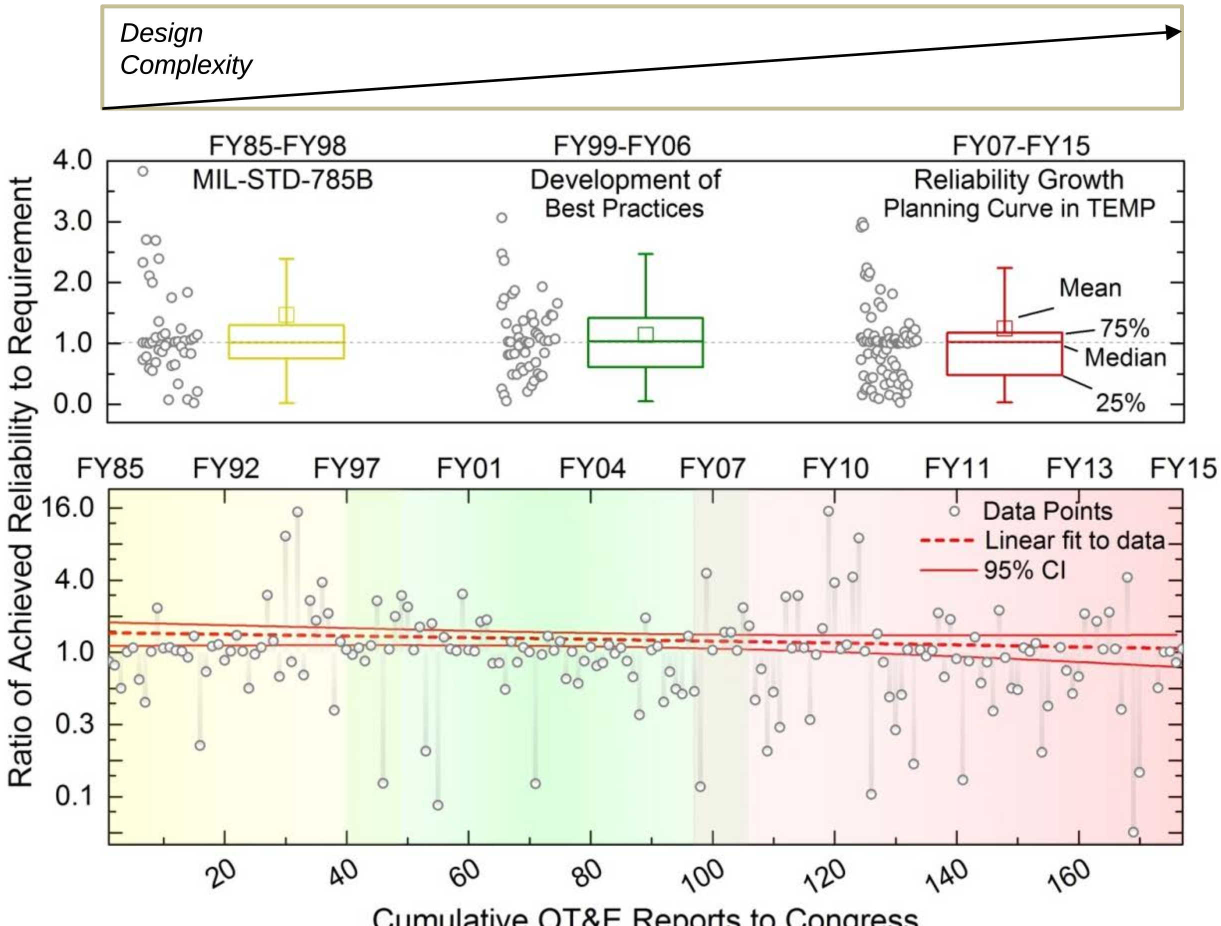
- Course briefings cover reliability planning and analysis activities that span the acquisition life cycle. Each briefing discusses review criteria relevant to DOT&E action officers based on DoD policies and lessons learned from previous oversight efforts

Time	Topic	Presenter
0900 – 0920	Course Introduction	Catherine Warner
0920 – 1000	RAM Requirements Review	Matthew Avery
1000 – 1045	Reliability Growth Planning	Jonathan Bell
1045 – 1100	Break	
1100 – 1145	Importance of Design Reviews in the Reliability Growth Planning Process	
1145 – 1245	Lunch Break	
1245 – 1330	TEMP Review and OT Planning	Rebecca Dickinson
1330 – 1400	Analysis of Reliability in DT	Rebecca Dickinson
1400 – 1415	Break	
1415 – 1530	Analysis of RAM data for LRIP/BLRIP reports	Matthew Avery

Motivation for Improving System Reliability

Why do it?

- Improve system reliability/meet thresholds
- Optimize test resources
- Improve system safety/suitability for user
- Reduce O&S Costs
- Quantify Risks
- Establish interim reliability goals



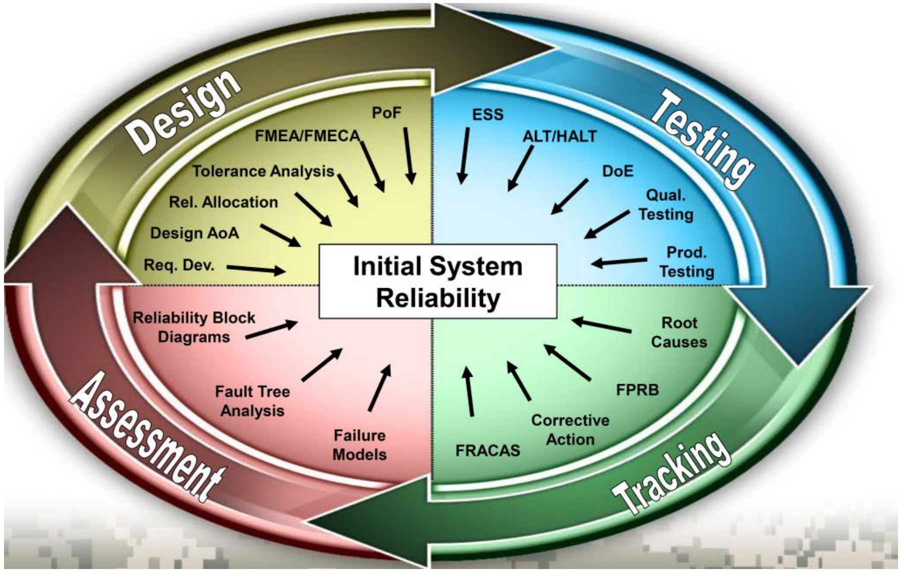
System Type	Fraction of Total Cost		
	RDT&E	Procurement	O&S
Ground Combat	4%	28%	68%
Rotary Wing	4%	31%	65%
Surface Ships	1%	39%	60%
Fighter Aircraft	5%	29%	66%

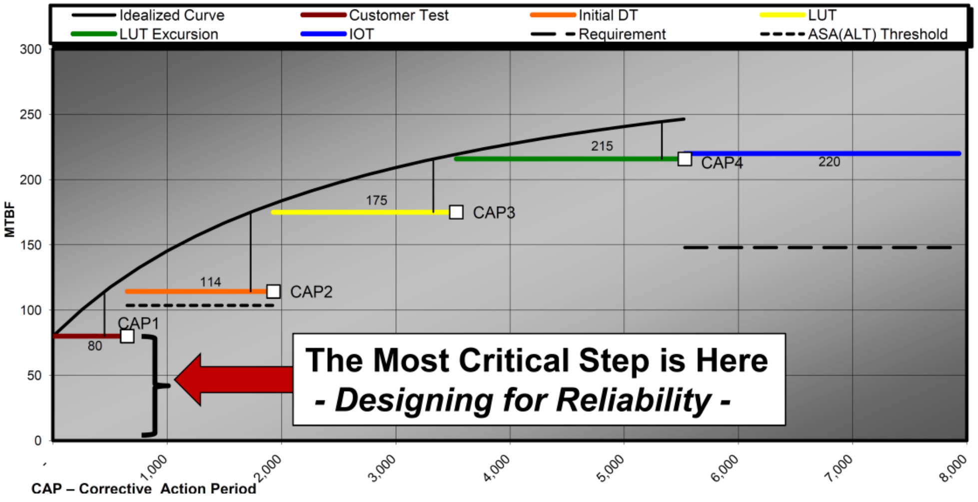
a. RDT&E – Research Development Test & Evaluation
b. O&S – Operations and sustainment
c. Data from AEC/AMSAA Reliability Course Notes," 21 Aug 2011.

Majority of cost here

a. CI – Confidence Interval *b. FY – Fiscal Year* *c. OT&E – Operational Test and Evaluation*

Design for Reliability (DfR)



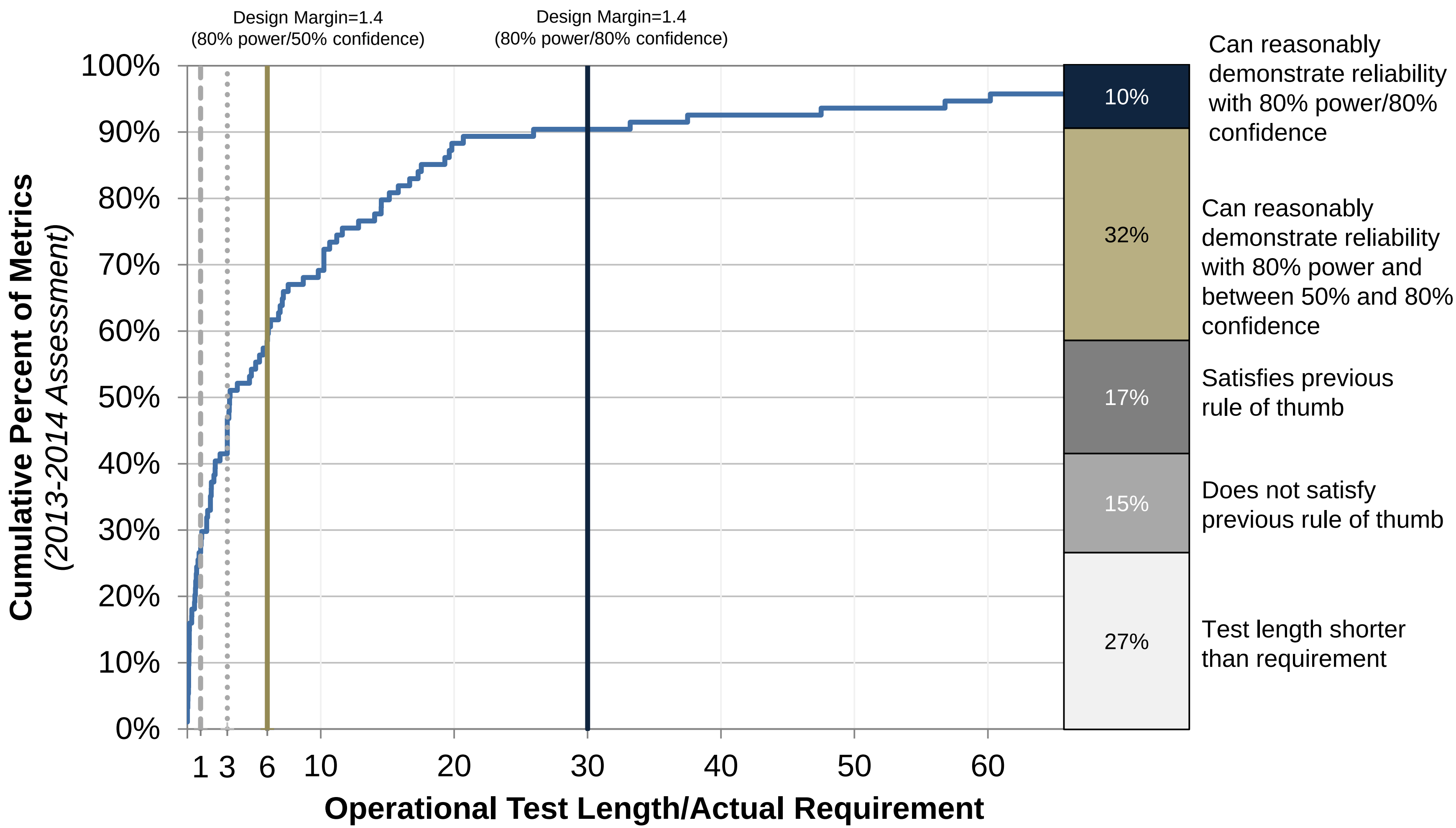


Reliability must be designed into the product from the beginning.

A common problem failure: to reach desired initial system reliability indicating failure in the design phase to engineer reliability into the system.

- Understand user requirements and constraints
- Design and redesign for reliability
- Produce reliable systems
- Monitor and assess user reliability

Evaluation of Test Adequacy for Assessing Reliability



Various methods are being used to develop adequate reliability test lengths

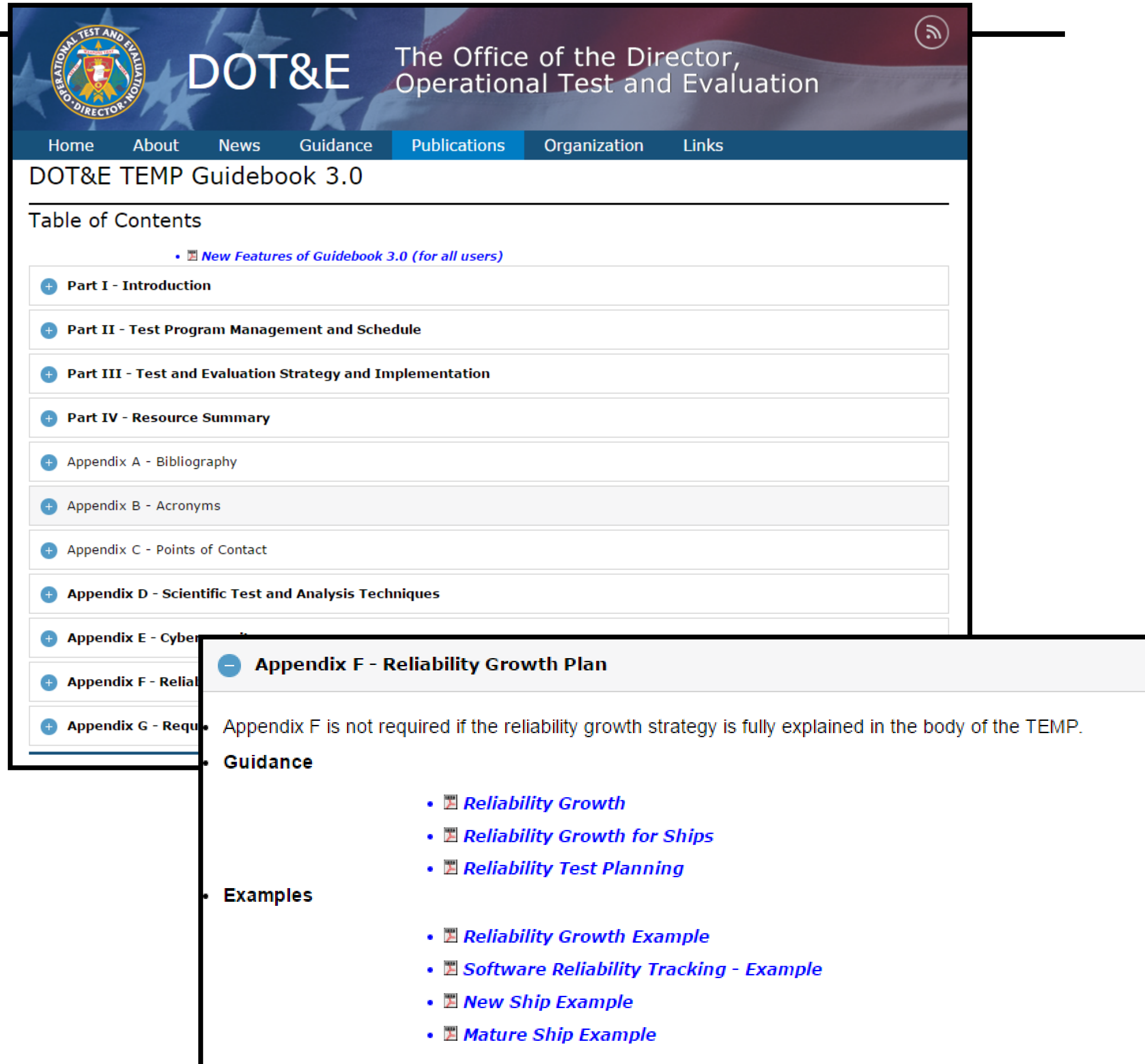
TEMP Guidebook 3.0 Reliability Updates

Reliability Growth Guidance

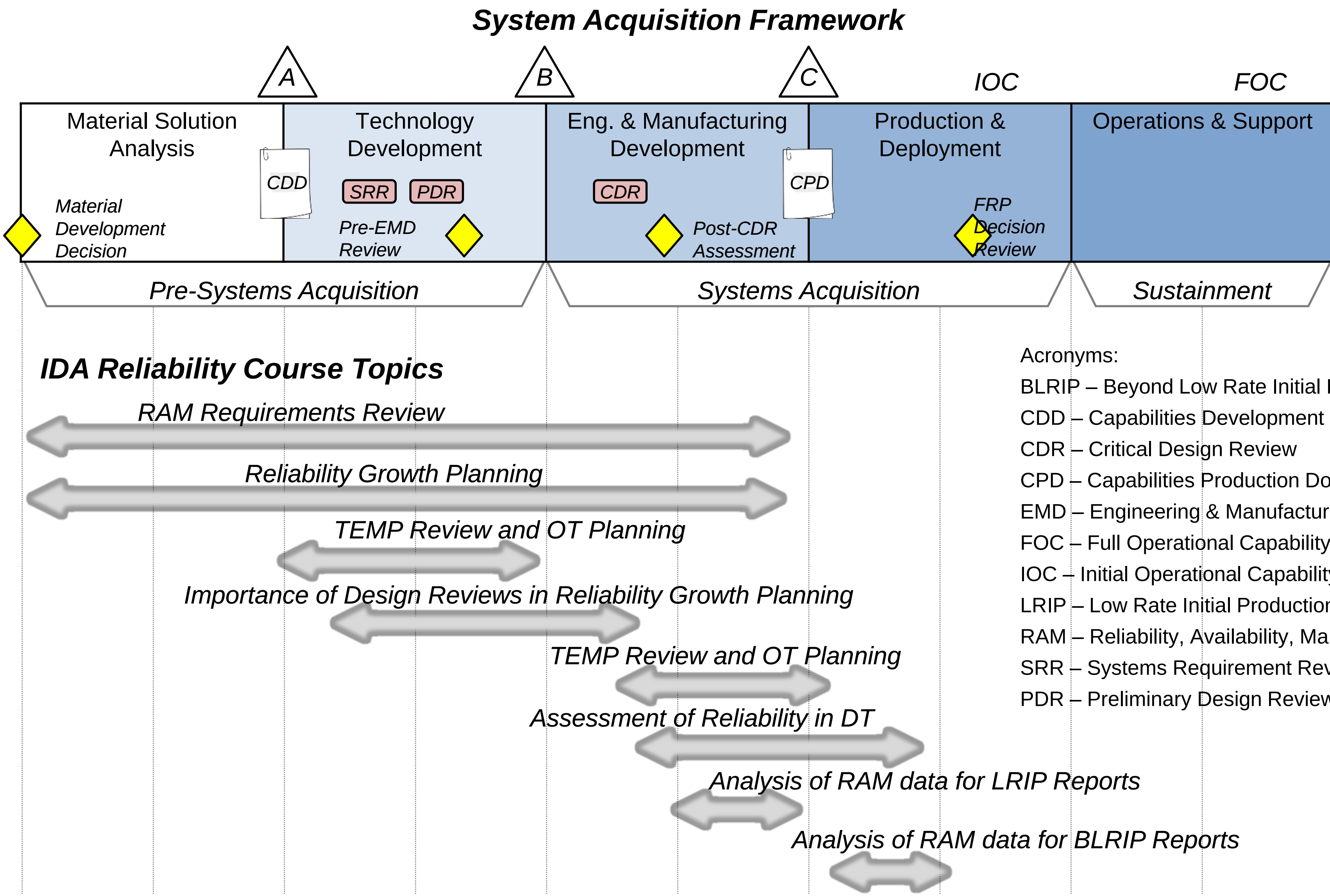
- Relatively unchanged from TEMP Guidebook 2.1

Reliability Test Planning Guidance

- New section of the TEMP Guidebook
- Emphases the use of operating characteristic curves for planning operational tests
- Provides guidance on using data collected outside of an operational test for reliability assessments



Topics Covered



Acronyms:

BLRIP – Beyond Low Rate Initial Production

CDD – Capabilities Development Document

CDR – Critical Design Review

CPD – Capabilities Production Document

EMD – Engineering & Manufacturing Development

FOC – Full Operational Capability

IOC – Initial Operational Capability

LRIP – Low Rate Initial Production

RAM – Reliability, Availability, Maintainability

SRR – Systems Requirement Review

PDR – Preliminary Design Review

Topics Covered (cont.)

Topic	Briefing Purpose/Objectives
Reliability, Availability, Maintainability (RAM) Requirements Review	• Highlight the importance of reviewing RAM requirements early in the program's lifecycle • Discuss criteria that should be considered during the review process
Reliability Growth Planning	• Provide an overview of the importance and process of reliability growth planning, focusing on information essential to support review of TEMPs and test plans • Demonstrate how to use the Projection Methodology (PM2) and Crow Extended reliability growth models
Importance of Design Reviews in the Reliability Growth Planning Process	• Highlight the importance of design reviews in the Reliability Growth Planning process, and identify the relevant questions to consider during design reviews • Provide programmatic examples of this process.
TEMP Review and Operational Test (OT) Planning	• Using examples, discuss how programs should document their reliability growth plan in the TEMP • Discuss criteria that should be considered during the review process • Describe how to assess the adequacy of an OT to evaluate reliability
Analysis of Reliability in Developmental Testing (DT)	• Explain how to determine if the proposed DT will be adequate to growth reliability • Provide an overview of DT activities that are essential to support reliability assessment and tracking
Analysis of RAM data for LRIP/BLRIP reports	• Discuss common methods for analyzing OT RAM data including development of confidence bounds, analysis of censored data, comparison to baseline/legacy, estimation of the reliability growth potential, subsystem failure analysis, etc.
Software Reliability	• Describe how the procedures for reliability growth planning and evaluation of software systems differ compared to hardware-based systems

Institute for Defense Analyses
4850 Mark Center Drive • Alexandria, Virginia 22311-1882

Reliability, Availability, Maintainability (RAM) Requirements Review

Matthew Avery
Laura Freeman

12 November 2015

Reliability in the DoD Context

Reliability *: the ability of an item to perform* ***a required function****, under given* ***environmental*** *and* ***operating conditions*** *and for a stated* ***period of time***
(ISO 8402, International Standard: Quality Vocabulary, 1986)

- **Operational mission reliability**
  - Most complex defense systems serve more than one required function (e.g., ships may provide transportation, defense, self-protection, etc.)
  - Multiple operating environments: desert, littoral, mountain, etc.
  - Operating conditions vary depending on mission
  - Requirements typically specify a fixed time period

- **An additional consideration in operational mission reliability**
  - Diverse population of system operators: crew-caused failures are still failures.

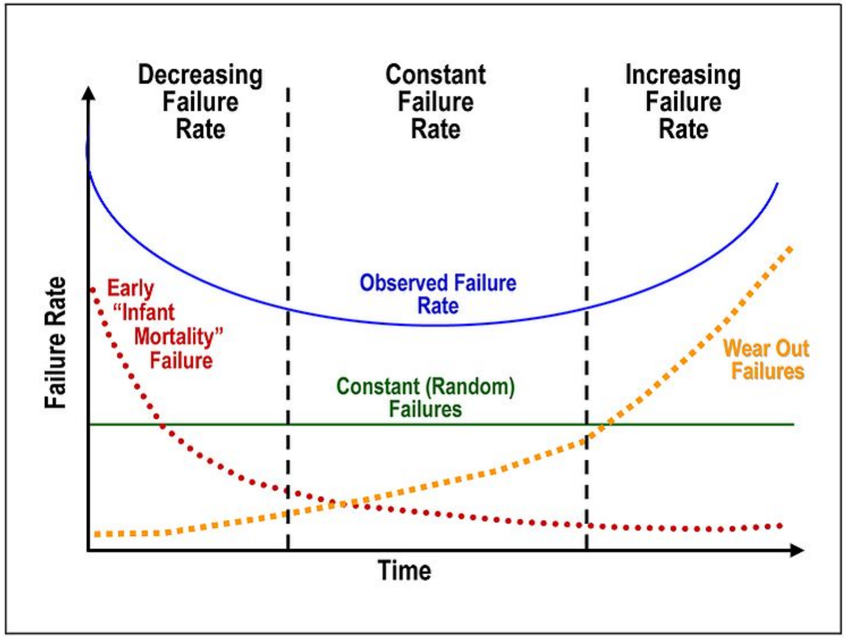
- **Concept of operations / Design reference mission – Essential for defining operational mission reliability**
  - Defines standard mission length
  - Provides a breakdown the expected activities during a mission
  - Can change over time as operational missions evolve

Scoring the Severity of a Failure

- **Operational Mission Failure (OMF) or System Abort (SA):** failure discovered during mission execution that result in an abort or termination of a mission in progress
    - » Reliability requirements are typically written in terms of OMFs or Sas.
- **Essential Function Failures (EFF) or Essential Maintenance Action (EMA):** failures of mission essential components. By definition all OMFs are EFFs
    - » EFFs include a large portion of the failure modes that drive maintenance costs and reduce system availability
- **Comparing EFFs and OMFs**
    - » Engine: temporary power failure vs. not starting at all
    - » Steering: excessive pulling in one direction vs. vehicle rolling
    - » Brakes: brake fluid leak/line worn vs. brake lock up

Traditional Reliability Analysis

- **A traditional reliability analysis models the mean time between operational mission failures (MTBOMF) as a constant value (constant failure rate)**
  - Exponential distribution
  - Failure mode is ignored
  - Only operational test data is considered

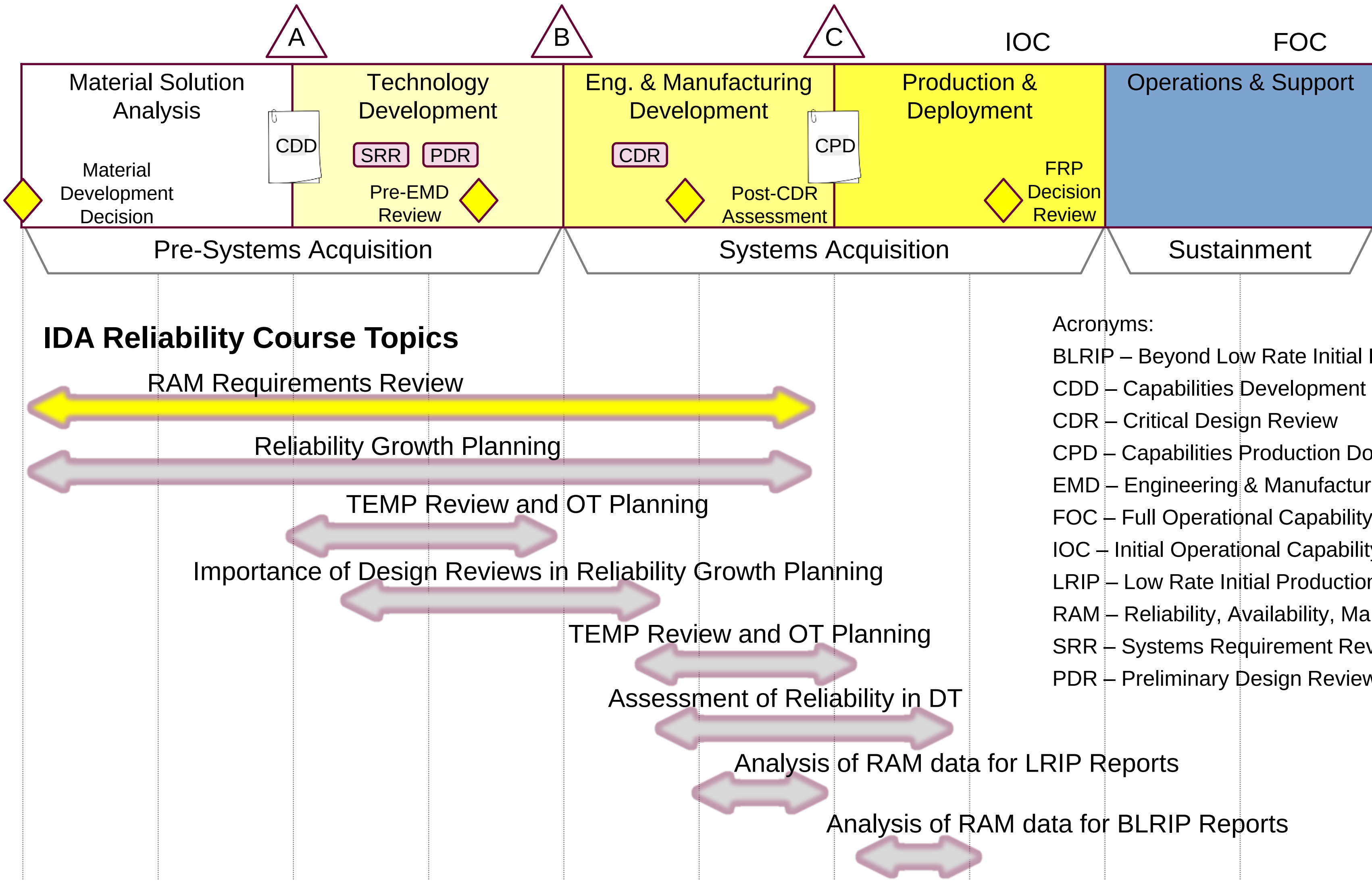


- **Reliability is calculated by:**

$$\widehat{\text{MMBOMF}} = \frac{\text{Total Time}}{\#\text{ of } \textbf{OMF}\text{ Failures}}$$

Timeline

System Acquisition Framework



Acronyms:

BLRIP – Beyond Low Rate Initial Production

CDD – Capabilities Development Document

CDR – Critical Design Review

CPD – Capabilities Production Document

EMD – Engineering & Manufacturing Development

FOC – Full Operational Capability

IOC – Initial Operational Capability

LRIP – Low Rate Initial Production

RAM – Reliability, Availability, Maintainability

SRR – Systems Requirement Review

PDR – Preliminary Design Review

Topics Covered

- **Importance of reviewing Reliability, Availability, Maintainability (RAM) requirements early in the program's lifecycle**
- **Criteria that should be considered when reviewing RAM requirements:**
  - What are your RAM requirements?
    - » Reliability, Availability, Maintainability Requirements
    - » By System Type (Single-Use, Repairable, One-off)
  - Levels of Failure
    - » Aborts or Operational Mission Failures
    - » Failures or Essential Function Failures
    - » Non Essential Function Failures
  - Mission-Level Reliability
  - Requirements in the Mission Context
  - Achievability of Requirements
  - Assessing the Failure Definition Scoring Criteria (FDSC) and/or Joint Reliability & Maintainability Evaluation Team (JRMET) documents

Importance of Early Review of RAM requirements

- **Requirements are generally established early in the program's lifecycle**
  - Before Milestone B for most programs
- **The first step in acquiring reliable systems is ensuring that they have achievable, testable, and operationally meaningful reliability requirements**
- **All systems have requirements**
  - Is this requirement operationally meaningful?
  - Is this requirement achievable?
    - » How reliable are similar systems that have already been fielded?
    - » Is the requirement achievable given its reliability growth plan?
  - Is the requirement testable?
- **Requirements Rationale in TEMP**
  - Starting at MS A
  - Reliability, maintainability, availability requirements should be addressed if not adequately addressed in the requirements document
  - When requirements are provided for all three metrics, DOT&E AO's should review to ensure they are mathematically consistent

Different Types of Systems

- **Single-use systems**
  - System is destroyed upon use
  - Missiles, rockets, MALD, etc.
  - Reliability is a simple probability (e.g., “Probability kill > 90%”)

- **Repairable Systems**
  - If the system breaks, it will be repaired and usage resumed
  - Tanks, vehicles, ships, aircraft, etc.
  - Reliability is typically time between events, i.e., failures, critical failures, aborts, etc.
    - » A howitzer must have a 75 percent probability of completing an 18-hour mission without failure.
    - » A howitzer mean time between failures must exceed 62.5 hours.

- **One-off systems**
  - Only a single (or very few) systems will be produced
  - Satellites, aircraft carriers, etc.
  - Like a repairable system, though often very few chances to improve reliability once system has been produced
  - Often no assembly line leading to different reliability concerns

Translating Reliability Requirements

Radar Program X's Capabilities Development Document (CDD):

After review, CDD determined that a clarification of the Mean Time Between Operational Mission Failure (MTBOMF) Key system Attribute (KSA) is appropriate and is rewritten as follows: "Radar Program X shall have a MTBOMF that supports a 90% probability of successful completion of a 24 Hour operational period (Threshold), 90% probability of successful completion of a 72 Hour operational period (Objective) to achieve the Operational Availability (Ao) of 90%"

- **90% probability of no Operational Mission Failure (OMF) over 24 hours**
  - Alternatively: Probability that time to failure > 24 hours is at least 90%

  → What is the average "Time to Failure"?
  → What is the distribution of failure times?

  - Based on exponential failure times:

$$P\left(T_{failure} > 24\right) = 0.9$$

$$\rightarrow MTBOMF \geq -24/\log(0.9) = 228\ hours$$

Translating Requirements: A Point of Caution

- **Assumptions in translation**
  - Mean is an appropriate metric to describe the failure distribution
  - The failures are exponentially distributed and therefore the failure rate is constant
  - No degradation ("wear-out") over time

- **Translation should be operationally meaningful**

- **Extremely high probability requirements can result in untestable/unrealistic mean duration requirements**

Probability of Mission Completion / Mission Duration	Mean Time Between Failure (MTBF)
99% (2-hour mission)	199 Hours
95% (2-hour mission)	39 Hours
95% (4-hour mission)	78 Hours

Availability Requirements

"The UAS shall achieve an A0 of at least 80% at IOC [Initial Operational Capability]."

- **Availability is a crucial measure of system performance and in many cases, is directly related to reliability**
- **Sometimes, reliability requirements are derived from availability requirements**
  - May need to make assumptions about repair times

  80% availability given 1 hour MTTR → MTBF= 4 hours:

  $$A_0 = \frac{MTBF}{MTBF + MTTR} \rightarrow .8 = \frac{MTBF}{MTBF + 1} \rightarrow MTBF = 4$$

  - Should only use this approach if no other reliability requirements are provided
  - Does not account for concurrent repairs

It is important to check for consistency between availability and reliability requirements!

MTTR – Mean Time To Repair
UAS – Unmanned Aerial System

Common Formulations of Availability

- **Each service defines availability differently**
  - See Memorandum of Agreement for different definitions and explanations
- **Operational availability A_O is the percentage of time that a system is available to perform its mission.**

$$A_O = \frac{Uptime}{Uptime + Downtime} = \frac{\sum Uptime_i}{\sum Uptime_i + \sum Downtime_i}$$

- A_O **is commonly computed:** $A_0 = \frac{MTBF}{MTBF+MTTR}$
- **Confidence interval methods for A_0 are equally valid for operational dependability D_O:**

$$D_O = \frac{MTBCF}{MTBCF + MTTRF}$$

- **Alternative formulation of $\boldsymbol{A_O}$:**

$$A_O = \frac{OT + ST}{OT + ST + TCM + TPM + TALDT}$$

MTBF- *Mean Time Between Failure*
MTTR- *Mean Time To Repair*
MTCBF- *Mean Time Between Critical Failure*
MTTRF- *Mean Time To Restore Function*
OT- *Operating Time*
ST- *Standby Time*
TCM- *Total Corrective Maintenance*
TPM- *Total Preventative Maintenance*
TALDT- *Total Administrative and Logistics Downtime*

Maintainability Requirements

"The UAS equipment and hardware components shall have a Mean Time to Repair (MTTR) for hardware of 1 hour."

- **Maintainability requirements often stated in terms of repair times ("mean time to repair" or "maximum time to repair")**
  - Some systems don't have specific values beyond being able to conduct field repairs

  "The Light Armored Vehicle-Recovery (LAV-R) shall enable the maintenance team to conduct battle damage repair and recovery."

- **Sometimes stated in terms of maintenance ratio**
  - "The Ground Combat Vehicle (GCV) will have a field level maintenance ratio (MR) that includes scheduled, unscheduled, and condition-based maintenance not to exceed 0.13 (Threshold) / 0.05 (Objective) maintenance man-hours per operating hour (MMH/OH)."

- **Median values and high percentile requirement can be more meaningful for systems with highly skewed repair times**
  - E.g., 90% of failures should be corrected within 5 hours
  - Or, the median repair for hardware should be 1 hour

Medians and percentiles are better maintainability requirements than means.

Non-standard Reliability Requirements

- **Effective Time On Station**
  - UAS: "The system must be sufficiently reliable and maintainable to achieve an Effective Time on Station (ETOS) rate of 85%."
    - » How do we define "Time On Station"?
    - » How do we treat pre-flight failures?
- **Littoral Combat Ship**
  - Capability Development Document (CDD) specifies target reliability for core mission as 0.8 in 720 hours
  - Four critical subsystems
    - » Total Ship Computing Environment (full-time)
    - » Sea Sensors and Controls (underway)
    - » Communications (full-time)
    - » Sea Engagement Weapons (on-demand)
  - System is "in series"
    - » System is up only if all critical subsystems are up

The requirements previously described may not apply to your system. In all cases it is important to understand the operational context of the requirement – If not documented in the requirements document, it should be in the TEMP.

Understanding Reliability in the Mission Context

DOT&E's decision for whether a system is Reliable is <u>not</u> dictated by the system's requirements

- **Identify the rationale for the reliability requirements and evaluate system reliability based on this rationale**

- **Understand the mission-level impact of reliability failures**
  - Most crucial systems/subsystems
  - Failure modes that have caused similar systems trouble in the past
  - Emphasis should be on completing the <u>*mission*</u> not the mean time between failures by themselves

- **Seek Contract/Requirement Documents for context**
  - Capability Production Document (CPD)
  - Capability Development Document (CDD)
  - Letters of clarification

Achievability of Requirements

- **Critical question: Are this system's reliability requirements achievable?**
  - Reliability for similar existing systems
  - Systems engineering plans

- **When requirements are unreasonable, push for an update early**
  - Unreasonable given existing technology
  - Unnecessary given mission
  - Untestable/unverifiable
    - » What is testable?

- **What is on contract?**
  - Typically, you will get what you pay for (or less!)
  - Identifying what is on contract will help you assess systems risk for achieving reliability requirement

- **Example of a high-risk reliability requirement:**
  - Early in the development of a tactical vehicle, the reliability requirement was set at 6,600 miles Mean Miles Between Operational Mission Failures (MMBOMF)
  - The legacy system being replaced achieved a reliability of ~1,200 miles MMBOMF
  - The tactical vehicle program eventually reduced the requirement to 2,400 miles MMBOMF

Scoring Reliability Testing

- **Failure Definition Scoring Criteria (FDSC)**
  - Master document describing failure modes and criteria for determining the level of a failure
  - Areas of concern/confusion should be addressed as early as possible and prior to testing
- **Joint Reliability and Maintainability Evaluation Team (JRMET) and Scoring Conferences**
  - May include representatives from Program Manager, Operational Test Agencies, and DOT&E
  - Events are scored by the JRMET at scoring conferences
  - Determine if a Test Incident Report is a failure and if so, how sever of a failure
  - Without a clearly discussed FDSC, reaching agreements may be difficult

Disagreements about reliability scoring criteria should be discussed ***prior*** *to the start of testing*

Failure Definition Scoring Criteria

- **The Failure Definition/Scoring Criteria (FDSC) is essential for defining failure, and scoring test results**
- **Failure Definitions**
  - Defines mission essential functions – minimum operational tasks the system must perform to accomplish assigned mission
    - » E.g., Maintain constant communications for a command and control system
- **Scoring Criteria**
  - Provides classification criteria that will be consistent across all phases of testing
    - » System Abort/ Operational Mission Failure
    - » Essential Function Failure/ Essential Maintenance Action
    - » Non Essential Function Failure/ Unscheduled Maintenance Action
    - » "No Test"
  - Rates Hazard/Severity of the failure or incident
  - Specifies chargeability of the failure
    - » Hardware, software
    - » Operator error
    - » Government furnished equipment (GFE)

DOT&E requires independent scoring of reliability failures – FDSC should provide guidance only! – 05 October 2012 Guidance Memo

Value of Lower Level Reliability Requirements

- **Examples of lower level reliability requirements**
  - Essential Function Failures (EFFs)
  - Unscheduled Maintenance Actions (UMAs)
- **Focus on maintenance burden of the system/system availability/logistical supportability of system/ensuring full mission capability**
- **More useful for measuring and tracking reliability growth**
- **More accurate estimates of system reliability**

Action Officers should encourage the use of lower level reliability requirements for systems with extremely high mission level requirements and/or systems with built-in redundancy.

Example Program: UAS Reliability Requirements

- **System of systems**
  - Modern systems are often complex and involve multiple subsystems
  - UAS includes 5 Air Vehicle, STUAS Recovery System, Launcher, and four Operator Work Stations
  - Government-Furnished Equipment (GFE) & Commercial Off-The-Shelf (COTS)

Notional System Configuration

UAS –Unmanned Aircraft System

Example Program: UAS Reliability Requirements

- **Air Vehicle reliability: MFHBA > 60 hours**
  - Five air vehicles in the system
- **Surface Components reliability: MTBA > 240 hours**
  - Launcher, Recovery System, Ground Control Station, etc.
  - Applies to both Land- and Ship-based configuration, though each configuration evaluated separately
- **Overall System Reliability: MFHBA > 50 hours**
- **Operational Availability > 80%**
  - Requires Recovery System, Launcher, at least 2 Air Vehicles, and at least two Operator Work Stations

Requirements include by subcomponent-level reliability and system-of-systems level reliability.

MFHBA – Mean Flight Hours Between Abort
MTBA – Mean Time Between Abort

Evaluating UAS Reliability Requirements

- **Are the requirements achievable?**
  - Other small Unmanned Aerial Vehicles (UAV) have achieved ~20 hours MFHBA

- **What is the impact of reliability in the mission context?**
  - 5 air vehicles in the system means considerable redundancy
    - » Pre-flight aborts to Air Vehicle (AV) may not impact system's ability to provide Intelligence, Surveillance, and Reconnaissance
  - Single points of failure for launcher and recovery system
    - » High reliability necessary for these systems

- **Avoid situational scoring**

Question: "Once the air vehicle is off station and RTB, do critical failures (e.g., AV crashes) count against MFHBA?"

Answer: ***YES!!!***

- Reliability calculations & reliability growth modeling assume constant failure → no situational scoring!

MFHBA – Mean Flight Hours Between Abort
RTB – Return To Base

Recommendations for AOs

- **Ensure reliability requirements are:**
  - Operationally meaningful – understand translations between mission completion
  - Testable
  - Achievable
- **Encourage the use of two-level reliability requirements**
  - Operational mission failures and essential function failures matter
- **Ensure consistency for reliability, maintainability, and availability requirements**
- **Participate in FDSC development**
- **Remember all failures count (GFE/Operator) and DOT&E scores independently**
  - Failure means system is not available
- **Avoid situational scoring**

Backup

Reliability Requirements

“The Amphibious Vehicle shall have a 0.77 probability of completing any single one of the scenarios described in the OMS/MP”

- **Testing the above is difficult as stated**
  - Would need to conduct many OMS/MP-sized missions to assess 77% success probability
- **Alternative framing**
  - “The Amphibious Vehicle shall have a Mean Time Between System Abort (MTBSA) of at least 69 hours”
    - » Based on an 18-hour OMS/MP
  - Based on success probability and length of OMS/MP (or “mission” or “flight”)
  - Easier to estimate MTBSA
  - Can translate MTBSA estimates back into “Probability of completing X-hour mission without failure”
    - » Useful for missions of variable length

OMS/MP - Operational Mode Summary/Mission Profile

Impact of Requirements on Future Testing

- **Will the requirements be testable as written?**
  - Ground Terminal Mean Time Between Critical Failure (MTBCF)
    - » Ground Fixed: 1150 hours
    - » Ground Transportable: 700 hours
- **Very high requirements necessitate very long tests**
  - One month's time: 720 hours
  - Three-failure test: 4300 hours

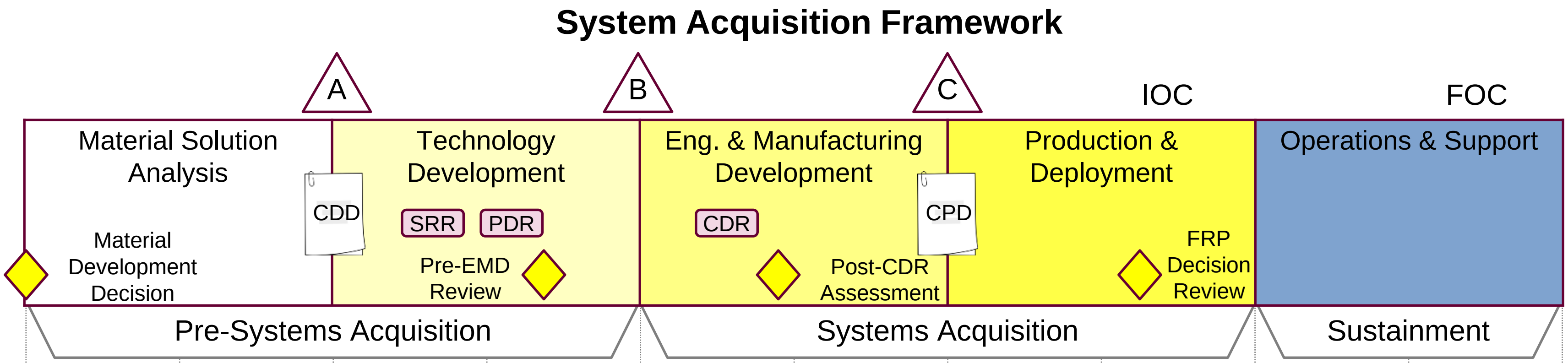
Some programs have especially high reliability requirements. In these cases, careful consideration must be given early on to the best approach to assess reliability.

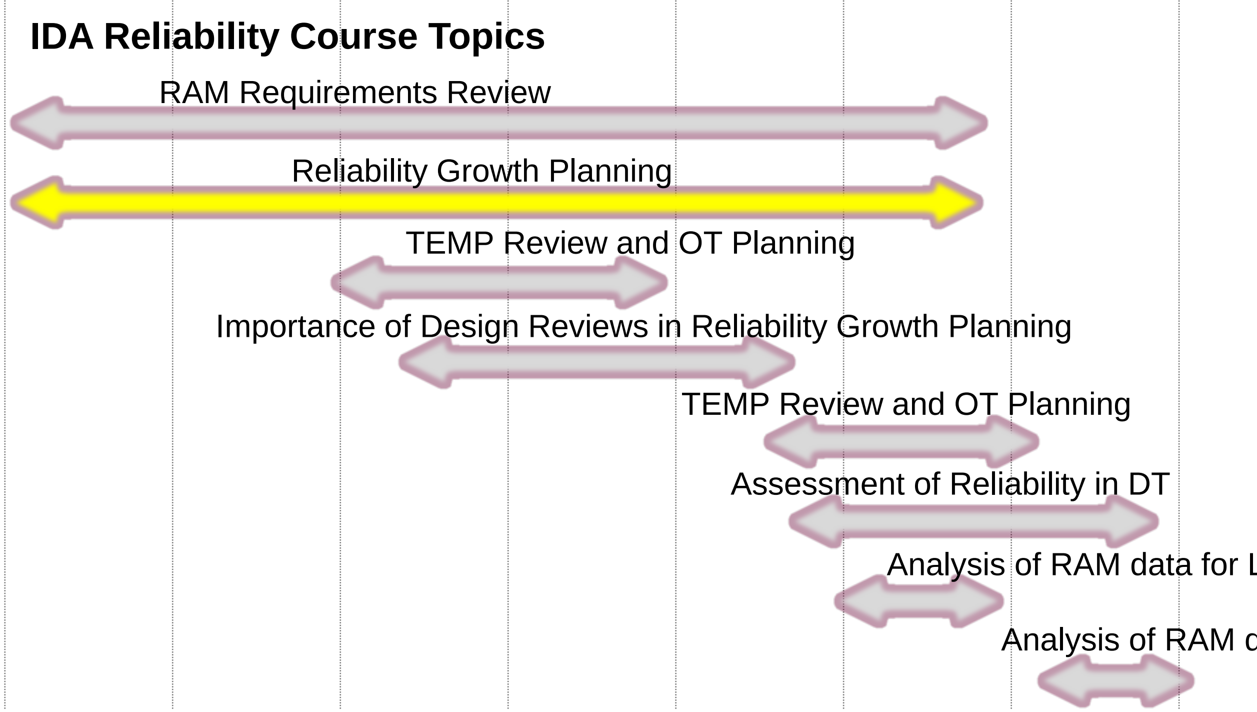
Institute for Defense Analyses
4850 Mark Center Drive • Alexandria, Virginia 22311-1882

Reliability Growth Planning

Jonathan L. Bell
2 June 2016

Timeline





Acronyms:

BLRIP – Beyond Low Rate Initial Production
CDD – Capabilities Development Document
CDR – Critical Design Review
CPD – Capabilities Production Document
EMD – Engineering & Manufacturing Development
FOC – Full Operational Capability
IOC – Initial Operational Capability
LRIP – Low Rate Initial Production
RAM – Reliability, Availability, Maintainability
SRR – Systems Requirement Review
PDR – Preliminary Design Review

Purpose of Briefing

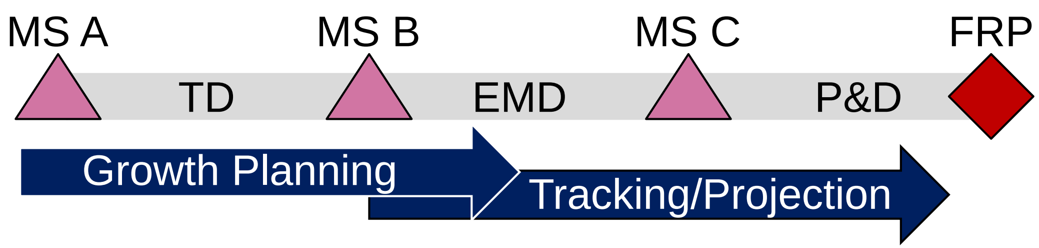
- Provide an overview that describes the importance and process of reliability growth planning, focusing on information essential to review of TEMPs and test plans

- Demonstrate how to use common reliability growth planning models
  - Planning Model Based on the Projection Methodology (PM2)
  - Crow Extended Reliability Growth Planning Models

Reliability Growth Planning Overview

- **Reliability Growth**
  - The process of eliminating initial design or manufacturing weaknesses in a system via failure mode discovery, analysis, and effective correction

- **Reliability Growth Planning is a structured process that is intended to occur early in the acquisition cycle**

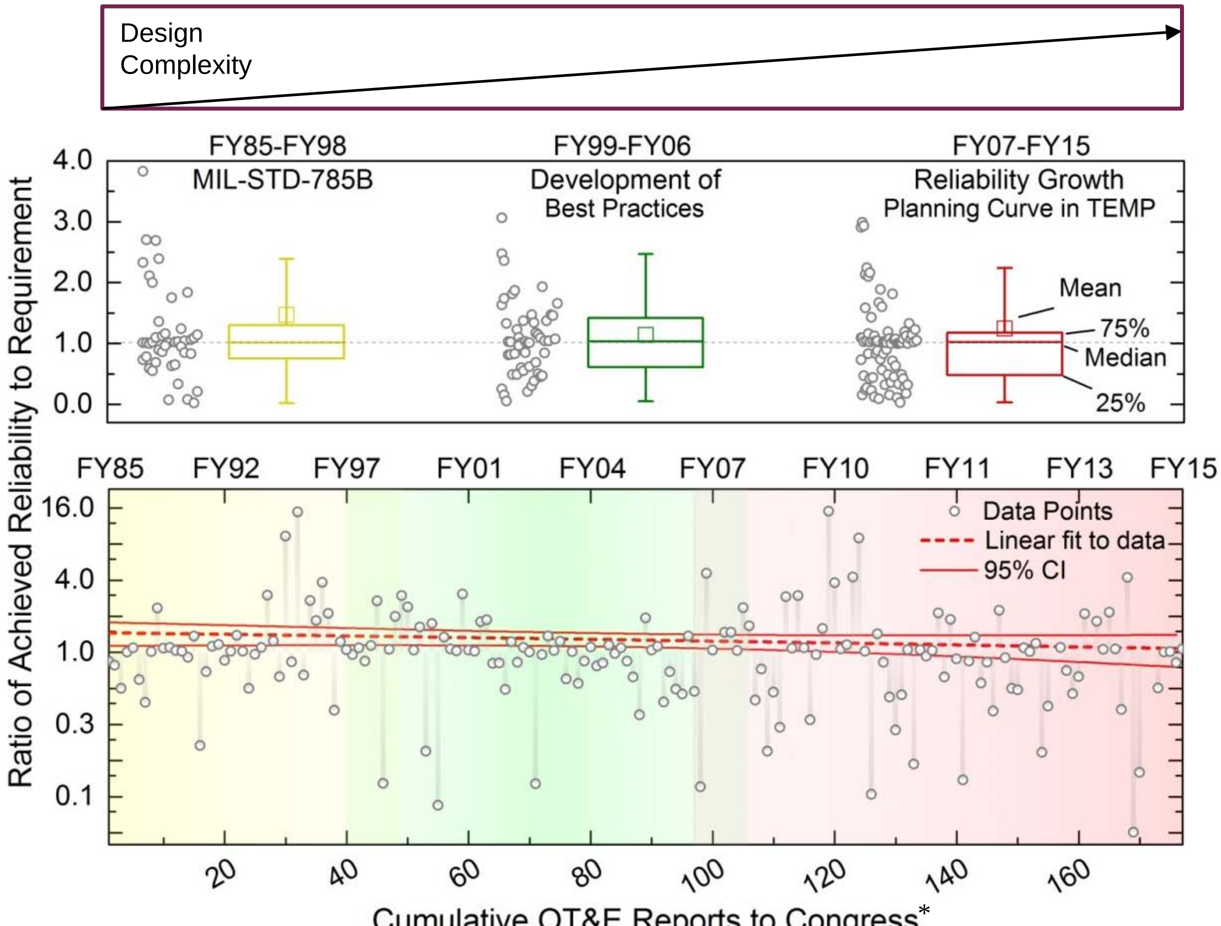


Acronyms: EMD – Engineering and Manufacturing Development; P&D – Production and Deployment; FRP – Full-Rate Production; TD – Technology Development; MS – Milestone

Motivation for Reliability Growth Planning

Why do it?

- Improve system reliability/meet thresholds
- Optimize test resources
- Improve system safety/suitability for user
- Reduce O&S Costs
- Quantify Risks
- Establish interim reliability goals



System Type	Fraction of Total Cost		
	RDT&E	Procurement	O&S
Ground Combat	4%	28%	68%
Rotary Wing	4%	31%	65%
Surface Ships	1%	39%	60%
Fighter Aircraft	5%	29%	66%

a. RDT&E – Research Development Test & Evaluation
b. O&S – Operations and sustainment
c. Data from AEC/AMSAA Reliability Course Notes," 21 Aug 2011.

Majority of cost here

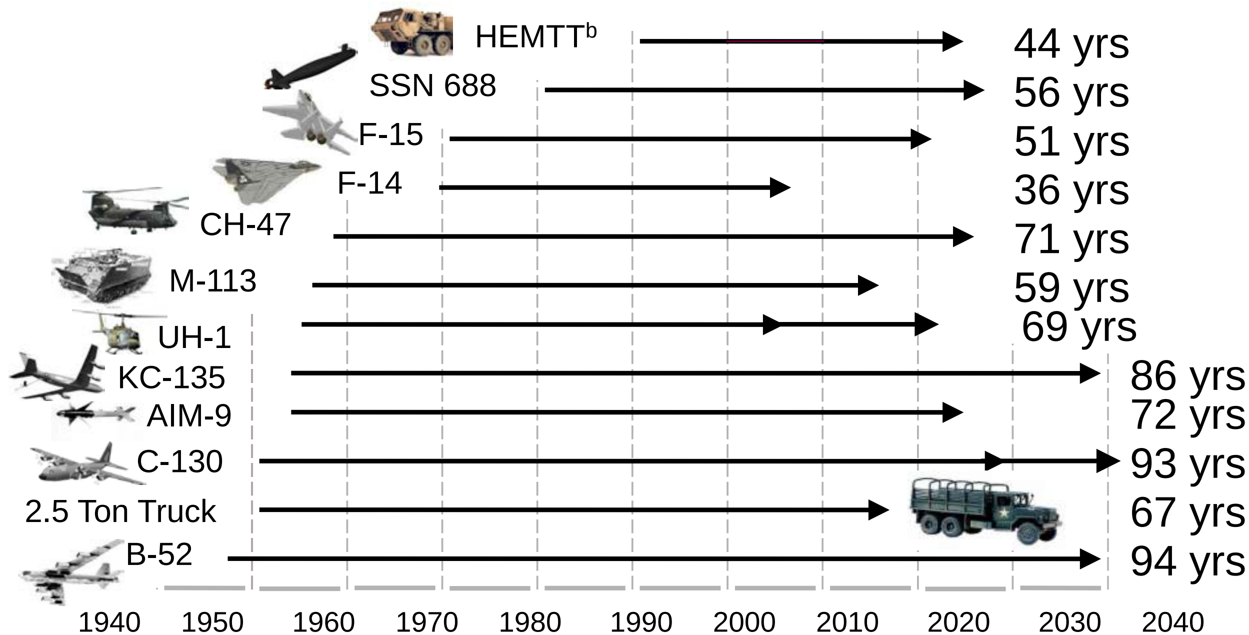
*Between 2000 and 2014, 22% of systems were found not reliable in IOT&E as result of failure modes attributed primarily to Software faults.

a. CI – Confidence Interval b. FY – Fiscal Year c. OT&E – Operational Test and Evaluation

Motivation for Reliability Growth Planning (cont.)

- **Systems are in service for a long time, which drives up Operations and Sustainment costs[a]**

  a. "Improving Reliability," Presentation to IDA by Dr. Ernest Seglie, 17 March 2009.
  b. HEMTT – Heavy Expanded Mobility Tactical Truck



- **It's Developmental Test and Evaluation's (DT&E) job, why should I do it?**
  - Some DOT&E oversight programs are not on DT&E oversight
  - Reliability growth planning is linked to entire acquisition cycle, including OT events
    - Part of reliability growth planning is ensuring that there is adequate testing/resources to evaluate reliability during OT
    - Data from a Limited User Test (LUT) or Operational Assessment (OA) is often analyzed to determine if system reliability is consistent with the reliability growth curve
    - Data from the Initial Operational Test and Evaluation (IOT&E) is often analyzed to prove whether system meets reliability requirements
    - The reliability growth contractual goal often depends on the length of the IOT&E

IDA

Reliability Growth Planning: DOT&E TEMP Guidebook 3.0

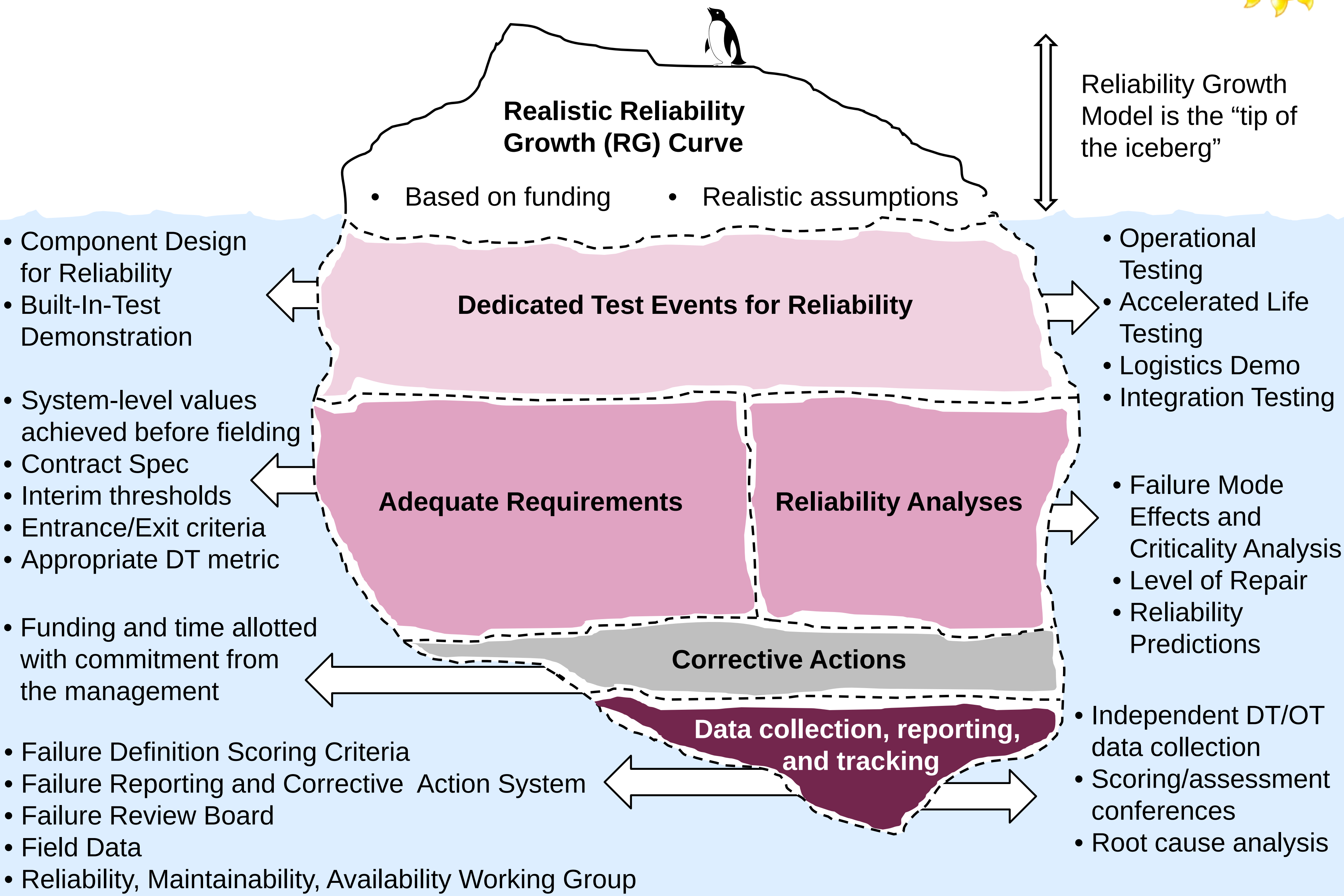
- **Includes additional specific guidance for different system types**
  - Software-intensive systems characterized by built-in redundancies that result in high reliability for the hardware (or hardware is not a component of the system), leaving the software reliability as the limiting factor (safety critical systems, automated information systems, and some space systems).
  - Hardware-only systems, which contain no software (bullets, personal protective equipment)
  - Hybrid systems containing a combination of software, hardware, and human interfaces. Critical functionality is a combination of hardware and software subsystems (complicated ground combat vehicles, aircraft, and ships) interfaces

- **For software-only systems, recommends:**
  - Addressing reliability growth by providing a reliability growth planning curve or a reliability growth tracking curve
  - Using the Crow-Extended Planning Model or the Planning Model based on Projection Methodology (PM2), if appropriate

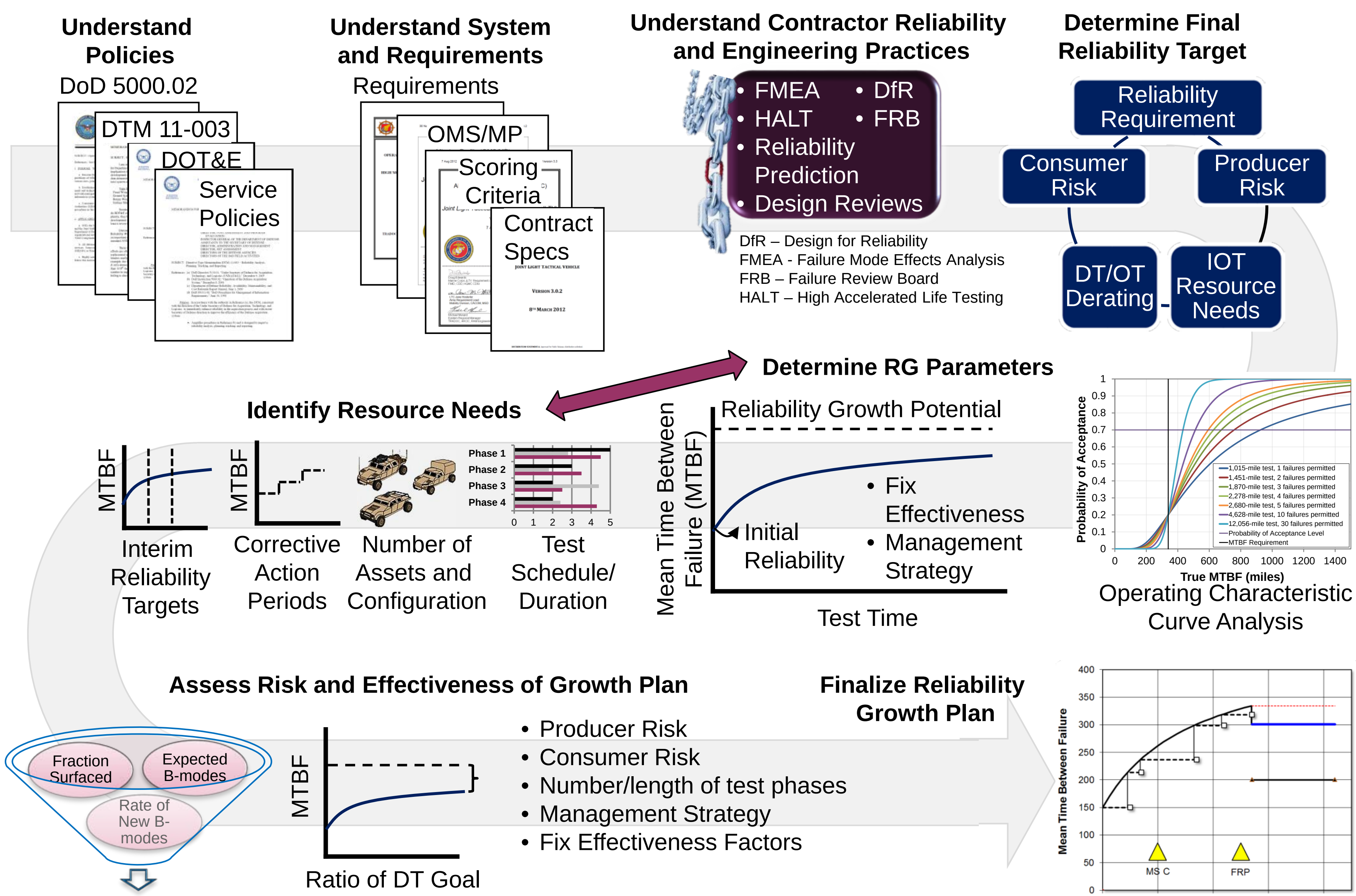
- **For hardware-only and hybrid systems, recommends :**
  - Developing reliability growth planning curves using PM2 Model or Crow-Extended Planning Model*

*PM2 and Crow Extended models encourage more realistic inputs that are based on the systems engineering and design process.

Elements of a Well-Run Reliability Growth Program



Reliability Growth (RG) Planning Process[a]



a. Figure adapted from ATEC Presentation on RG Planning, Joint Service RAM WG Meeting, SURVICE Engineering, Aberdeen, MD, 10-13 Jan 2011

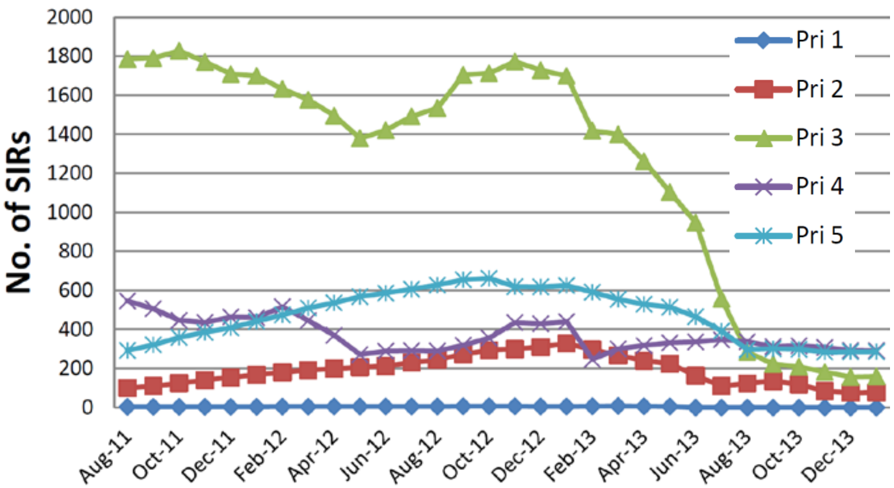
Reliability Growth (RG) Planning Process (cont.)

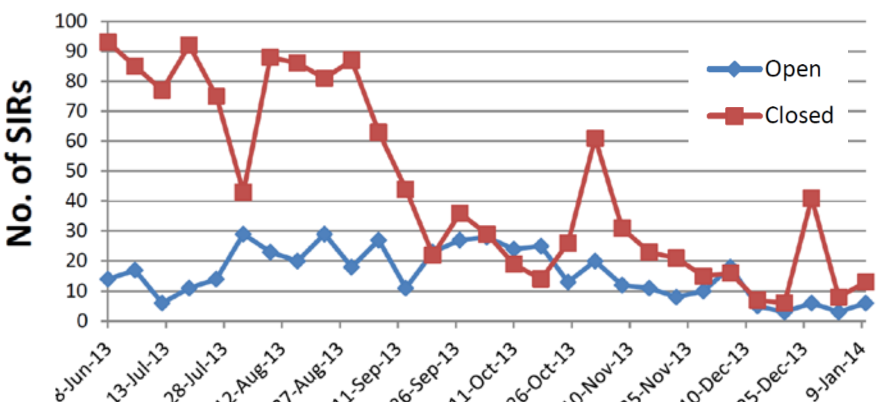
- **Reliability Growth Planning for Software Intensive Systems**
  - Follows a similar process as planning for hybrid and hardware-only systems:
    - Requires robust systems engineering support, dedicated testing, adequate funding and schedule time, reasonable requirements, scoring criteria, data collection and reporting, meetings to assess and score data, etc.
    - Ideally, should have an OT of sufficient length to demonstrate compliance with requirement
    - Can be described using Non-Homogeneous Poisson Process (NHPP) models in the relation to time (e.g., the AMSAA PM2 and Crow Extended Models) due to their simplicity, convenience, and tractability.
  - Growth planning can also be accomplished using a reliability tracking curve
    - IEEE Standard 1633 describes the practice for software reliability prediction prior to testing
    - Typically involves tracking the number of open and resolved problem reports over time
  - The basis for scoring criteria and prioritization can be found in IEEE Standard 12207 for Systems and Software Engineering — Software Life Cycle Processes:

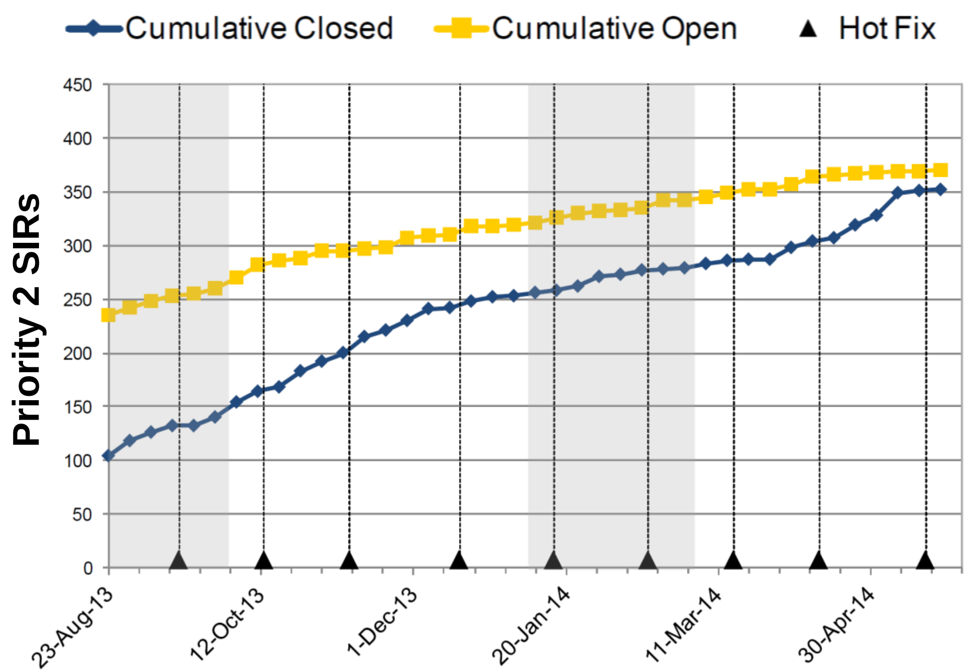
Priority	Applies if the Problem Could
1	Prevents the accomplishment of an essential capability, or jeopardizes safety, security, or requirement designated as critical
2	Adversely affects the accomplishment of an essential capability and no workaround solution is known, or adversely affects technical, cost, or schedule risks to the project or to life cycle support of the system, and no work-around solution is known
3	Adversely affects the accomplishment of an essential capability but a work-around solution is known, or adversely affects technical, cost, or schedule risks to the project or to life cycle support of the system, but a work-around solution is known
4	Results in user/operator inconvenience or annoyance but does not affect a required operational or mission essential capability, or results in inconvenience or annoyance for development or maintenance personnel, but does not prevent the accomplishment of those responsibilities
5	All other effects

Reliability Growth (RG) Planning Process (cont.)

- **Notional Examples of reliability tracking curves for Software Intensive Systems**



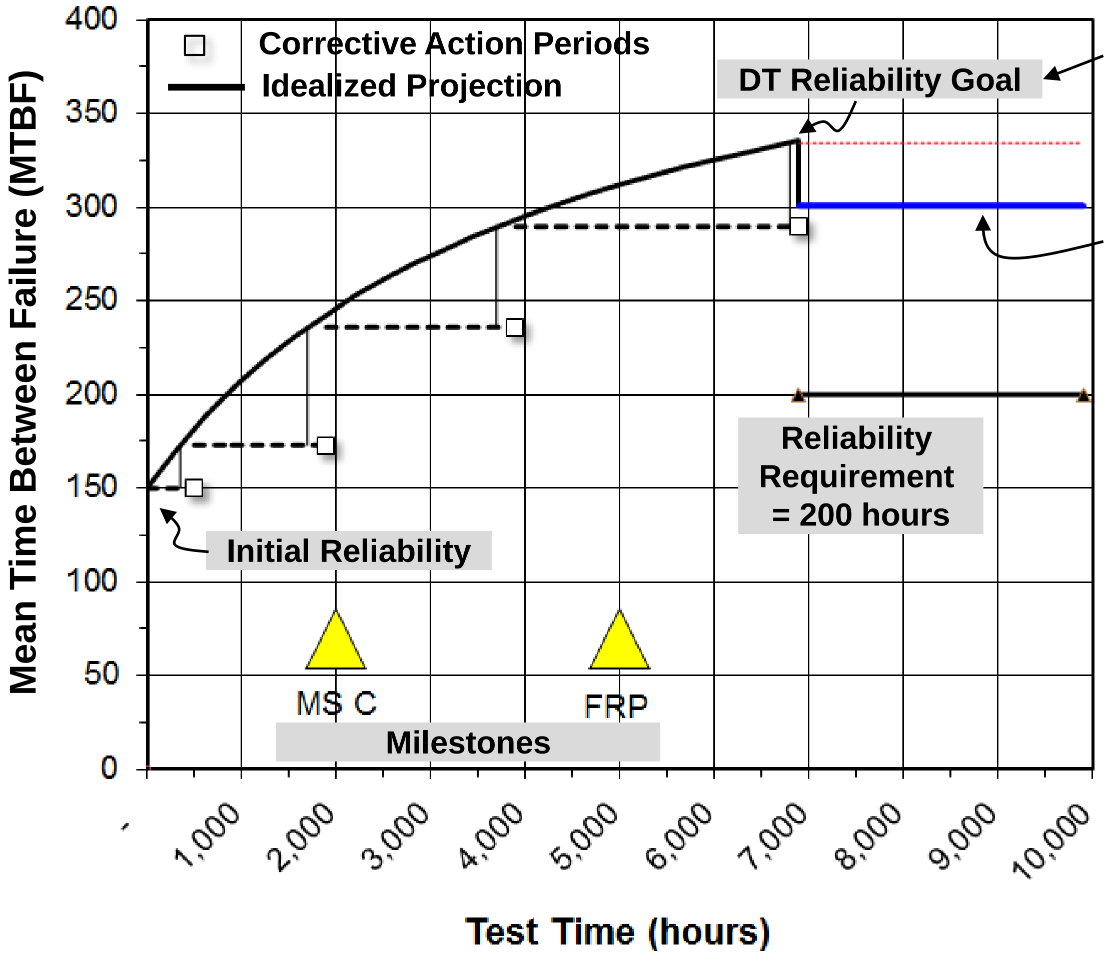




*SIR – Software Issue Report

Planning Model based on Projection Methodology (PM2)

- **Typical PM2 Reliability Growth Planning Curve**



Calculated by dividing OT reliability goal of 300-hrs MTBF by 0.9 to account for planned 10% reduction in DT MTBF due to OT environment

OT reliability goal of 300-hrs MTBF based on demonstrating 200-hr MTBF requirement with 20% consumer and 20% producer risks

Other Model Parameters

- **Management Strategy - fraction of the initial system failure intensity due to failure modes that would receive corrective action. Considers A and B modes, which are failure modes that will (B modes) or will not (A modes) be addressed via corrective action**
- **Average Fix Effectiveness Factor - the reduction in the failure rate due to implementation of a corrective actions**
- **Growth Potential - theoretical upper limit on reliability which corresponds to the reliability that would result if all B-modes were surfaced and fixed with the realized failure mode FEF values**

Acronyms: FRP – Full Rate Production
MS - Milestone

"Department of Defense Handbook Reliability Growth Management," MIL-HDBK-189C, 24 June 2011.

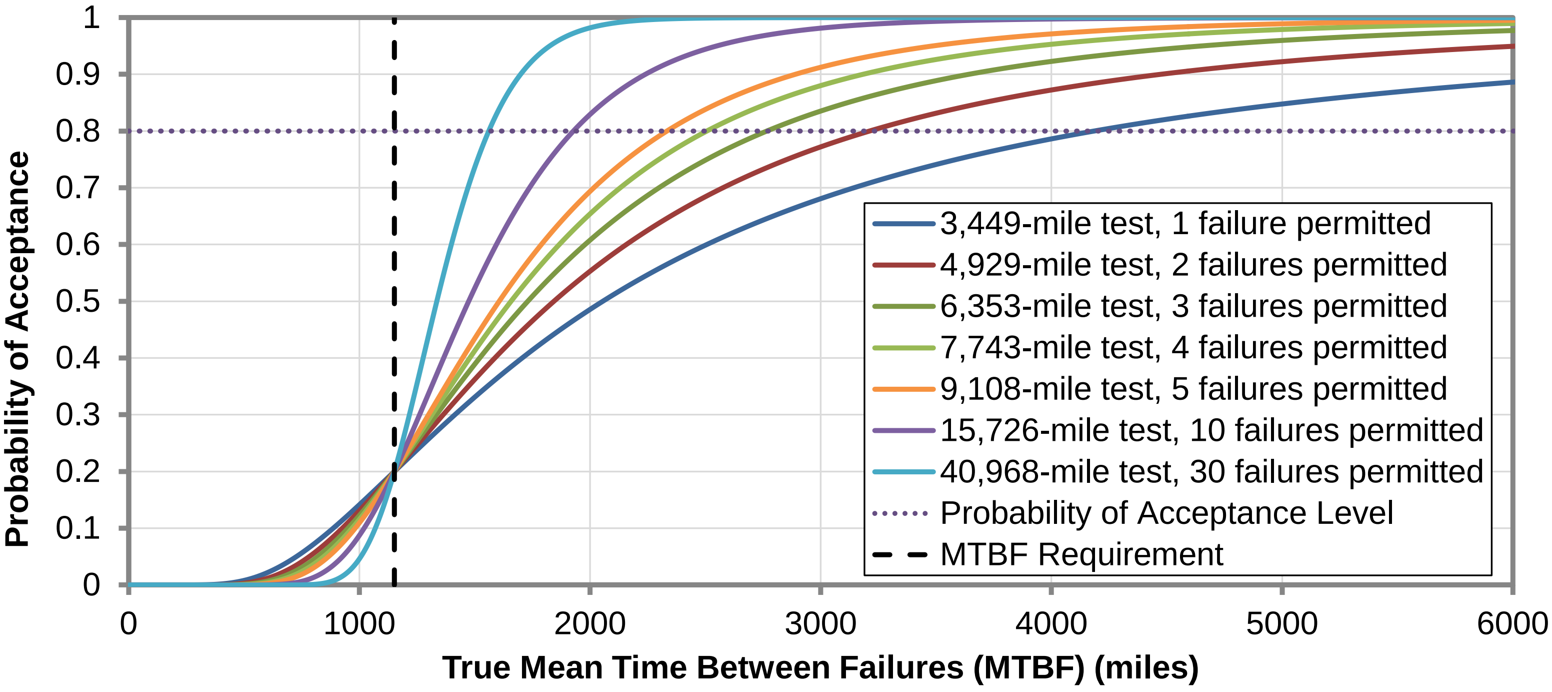
Planning Model based on Projection Methodology (PM2) (cont.)

Operating Characteristic (OC) Curves

- Help us to consider whether test scope is adequate to assess system reliability
- Illustrate allowable test risks (consumer's and producer's risks) for assessing the progress against the reliability requirement

Reliability requirement	1152
Confidence (1-consumer risk)	0.8
Probability of Acceptance (producer risk)	0.8
Ratio of DT reliability goal to requirement	1.75

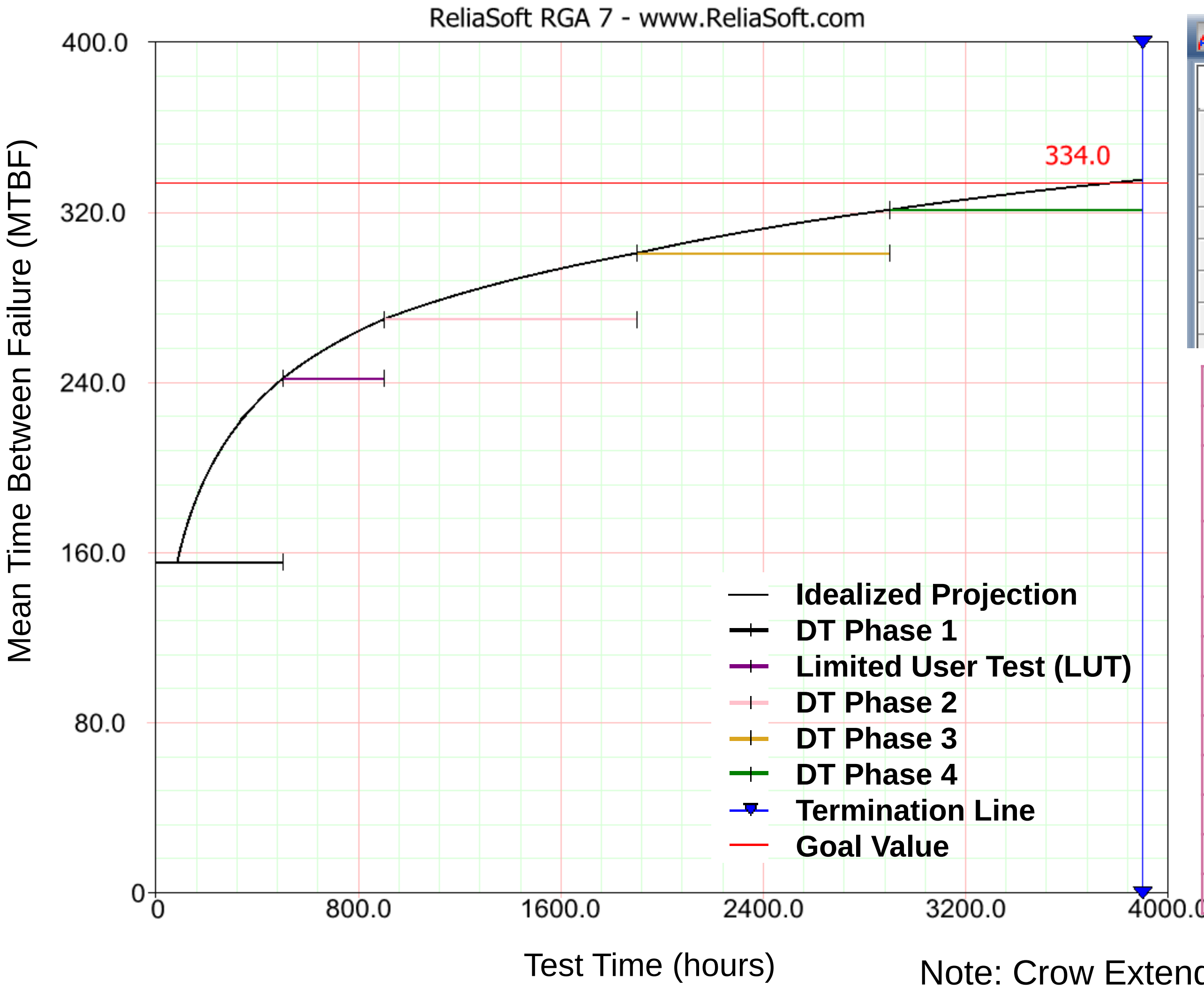
User inputs



In Class Exercise

Crow-Extended Reliability Growth Model

- **Typical Crow-Extended Reliability Growth Planning Curve**



Growth Planning: Plan1

A1 DT Phase 1

	Phase Name	Cumulative Phase Time	Average Fix Delay
1	DT Phase 1	500	150
2	LUT	900	0
3	DT Phase 2	1900	200
4	DT Phase 3	2900	200
5	DT Phase 4	3900	100

Planning Information	
Input	
Goal Mean Time Between Failure (MTBF)	334
Growth Potential Design Margin	1.39
Average Fix Effectiveness	0.70
Management Strategy	0.95
Discovery Beta	0.57
Results	
Initial Time [t(0)]	84
Initial MTBF	155
Final MTBF	336
Time at Goal	3,677

Note: Crow Extended does not use OC curves to determine the reliability growth goal.

Crow-Extended Reliability Growth Model (cont.)

In Class Exercise

Common Reasons Why Programs Fail to Reach Reliability Goals and What We Can Do About It

1. Failure to start on the reliability growth curve due to poor initial reliability of design
2. Failure to achieve sufficient reliability growth during developmental testing (DT)
3. Failure to demonstrate required reliability in operational testing (OT)

Failure to start on the reliability growth curve due to poor initial reliability of design

Common Causes	Recommended DoD Mitigations
Poor integration or lack of a "design for reliability" effort	Review contractor's reliability engineering processes; Establish contractual requirements that encourage system engineering "best practices"
Unrealistic initial reliability predictions based on MIL-HDBK-217	Review prediction methodology; Require/encourage more realistic prediction methods such as physics of failure method using validated models and/or test data; Have experts review contractor software architecture and specifications
Early contractor testing is carried out in a non-operational environment	Understand how the contractor conducted early testing; Encourage contractor to test system in an operationally realistic environment as early as possible
Unrealistic reliability goals relative to comparable systems or poorly stated requirements	Compare reliability goals to similar systems; Push for more realistic requirements
Overestimating the reliability of COTS/GOTS in a military environments	Communicate the operational environment to the contractor, and the contractor, in turn, has to communicate that information to any subcontractors; If available, consider field data and prior integration experience to estimate reliability
Lack of understanding of the definition of "system failure"	Review system design/scoring criteria early and ensure all parties understand and agree with it; Communicate scoring criteria in Request For Proposal
Reliability requirement is very high and would require impractically long tests to determine the initial reliability with statistical confidence	Consider using "lower-level" reliability measures (e.g., use MTBEFF, instead of MTBSA); Investigate if the specified level of reliability is really required for the mission; Emphasize the importance of having a significant design for reliability efforts

MTBEFF – Mean Time Between Essential Function Failures   MTBSA – Mean Time Between System Aborts

Common Reasons Why Programs Fail to Reach Reliability Goals and What We Can Do About It (cont.)

Failure to achieve sufficient reliability growth during developmental testing (DT)

Common Causes	Recommended Mitigation
Development of the reliability growth planning curve was a "paper exercise" that was never fully supported by funding, contractual support, and systems engineering activities	Verify reliability program is included in contracting documents and that there is sufficient funding to support testing and system engineering activities; Ensure program has processes in place to collect and assess reliability data; Investigate realism of reliability growth model inputs
Insufficient testing or time to analyze failure modes and devise/implement corrective actions	Evaluate how many B-mode failures are expected to surface over the test period; Ensure there are sufficient test assets and push for additional assets when the testing timeline is short; Evaluate if there will be sufficient time to understand the cause of failures and develop, implement, and verify corrective actions
Urgent fielding of systems that are not ready for deployment	
Inadequate tracking of software reliability or testing of patches	Ensure contract includes provisions to support software tracking and analysis; TEMP should define how software will be tracked/prioritized
System usage conditions or environment changed during testing	Analyze data to see if the failure mode distributions varied with changing conditions, Consider whether to reallocate resources and conduct additional testing in more challenging conditions
Initial design or manufacturing processes underwent major changes during testing	Discuss whether it is necessary to rebaseline the reliability growth planning curve based on the new design
System/subsystem components reaches wear-out state during testing	Investigate cause of wear-out; Consider recommending redesign for subsystems showing early wear-out or taking steps to mitigate overstresses to these components, if applicable
Reliability requirement is very high and would require impractically long tests to surface failure modes and grow reliability	Consider using "lower-level" reliability measures (e.g., use MTBEFF, instead of MTBSA); Investigate if the specified level of reliability is really required for the mission; Emphasize the importance of having a significant design for reliability efforts

MTBEFF – Mean Time Between Essential Function Failures MTBSA – Mean Time Between System Aborts

Common Reasons Why Programs Fail to Reach Reliability Goals and What We Can Do About It (cont.)

Failure to demonstrate required reliability in operational testing (OT)

Common Causes	Recommended Mitigation
Reliability of the system was poor coming in to the OT	Encourage program to establish OT reliability entrance criteria and ensure these criteria are achieved prior to entering the OT
User employment, environment, and/or system configuration was different in OT than in DT	Seek to operationalize reliability testing in DT to the maximum extent possible
Data collection and scoring procedures were different in OT compared to DT	Ensure data collection in DT and OT are adequate; Encourage program office and test agency to establish procedures that encourage data collection quality and consistency; Perform a pilot test to assess data collection adequacy
OT length was too short	Use operating characteristic curves and other appropriate statistical methods to scope the OT length; Use DT data to estimate system reliability

Takeaway Points

- Given the poor performance of producing reliable systems in the DoD, development of a comprehensive reliability growth plan is important and is required by policy

- Reliability planning is more than producing a growth curve; it requires adequate funding, schedule time, contractual and systems engineering support, reasonable requirements, scoring criteria, data collection and assessment, etc.

- Reliability growth planning models, such as PM2 and Crow-Extended, provide useful ways to quantify how efforts by the management can lead to improved reliability growth over time

- Reliability growth planning for software intensive systems generally follows a similar process as planning for hybrid and hardware-only systems, although use of a tracking curve can also support quantification of growth planning efforts

- Programs fail to reach their reliability goals for a variety of reasons; development of a robust growth plan early on can help avoid some of the common pitfalls

Reliability References

DOT&E references

- "DOT&E TEMP Guide," 28 May 2013 (Version 3.0 Update in progress)
- " Independent Operational Test and Evaluation (OT&E) Suitability Assessments," Memo, 5 Oct 2012.
- "State of Reliability," Memo from Dr. Gilmore to Principal Deputy Under Secretary of Defense (AT&L), 30 June 2010.
- "Next Steps to Improve Reliability," Memo from Dr. Gilmore to Principal Deputy Under Secretary of Defense (AT&L), 18 Dec 2009.
- "Test and Evaluation (T&E) Initiatives," Memo from Dr. Gilmore to DOT&E staff, 24 Nov 2009.
- "DOT&E Standard Operating Procedure for Assessment of Reliability Programs by DOT&E Action Officers," Memo from Dr. McQuery, 29 May 2009.
- "DoD Guide for Achieving Reliability, Availability, and Maintainability," DOT&E and USD(AT&L), 3 Aug 2005.

Other references

- "Reliability Growth: Enhancing Defense System Reliability," National Academies Press, 2015.
- "Department of Defense Handbook Reliability Growth Management," MIL-HDBK-189C, 14 June 2011.
- "Improving the Reliability of U.S. Army Systems," Memo from Assistant Secretary of the Army AT&L, 27 June 2011.
- "Reliability Analysis, Tracking, and Reporting," Directive-Type Memo from Mr. Kendall, 21 March 2011.
- "Department of Defense Reliability, Availability, Maintainability, and Cost Rationale Report Manual," 1 June 2009.
- "Implementation Guide for U.S. Army Reliability Policy," AEC, June 2009.
- "Reliability Program Standard for Systems Design, Development, and Manufacturing," GEIA-STD-009, Aug. 2008.
- "Reliability of U.S. Army Materiel Systems," Bolton Memo from Assistant Secretary of the Army AT&L, 06 Dec 2007.
- "Empirical Relationships Between Reliability Investments And Life-cycle Support Costs," LMI Consulting, June 2007.
- "Electronic Reliability Design Handbook," MIL-HDBK-338B, 1 Oct. 1998.
- "DoD Test and Evaluation of System Reliability, Availability, and Maintainability: A primer," March 1982.

Software

- AMSAA Reliability Growth Models, User Guides and Excel files can be obtained from AMSAA.
- RGA 7, Reliasoft.
- JMP, SAS Institute Inc.

Backup Slides

PM2 Continuous RG Curve Risk Assessment

Category	Low Risk	Medium Risk	High Risk
$\frac{\text{MTBF Goal (DT)}}{\text{MTBF GrowthPotential}}$	< 70%	70 - 80%	> 80%
IOT&E Producer's Risk	≤ 20%	20⁺ - 30%	> 30%
IOT&E Consumer's Risk	≤ 20%	20⁺ - 30%	> 30%
Management Strategy	< 90%	90 - 96%	> 96%
Fix Effectiveness Factor	≤ 70%	70⁺ - 80%	> 80%
$\frac{\text{MTBF Goal (DT)}}{\text{MTBF Initial}}$	< 2	2 - 3	> 3
Time to Incorporate and Validate Fixes in IOT&E Units Prior to Test	Adequate time and resources to have fixes implemented & verified with testing or strong engineering analysis	Time and resources for almost all fixes to be implemented & most verified w/ testing or strong engineering analysis	Many fixes not in place by IOT&E and limited fix verification

"AEC/AMSAA Reliability Short Course," SANGB, MI, 22 August 2012.

PM2 Continuous RG Curve Risk Assessment(cont.)

Category	Low Risk	Medium Risk	High Risk
Corrective Action Periods (CAPs)	5 or more CAPs which contain adequate calendar time to implement fixes prior to major milestones	3 - 4 CAPs but some may not provide adequate calendar time to implement all fixes	1- 2 CAPs of limited duration
Reliability Increases after CAPs	Moderate reliability increases after each CAP result in lower-risk curve that meets goals	Some CAPs show large jumps in reliability that may not be realized because of program constraints	Majority of reliability growth tied to one or a couple of very large jumps in the reliability growth curve
Percent of Initial Problem Mode Failure Intensity Surfaced	Growth appears reasonable (i.e. a small number of problem modes surfaced over the growth test do not constitute a large fraction of the initial problem mode failure intensity)	Growth appears somewhat inflated in that a small number of the problem modes surfaced constitute a moderately large fraction of the initial problem mode failure intensity	Growth appears artificially high with a small number of problem modes comprising a large fraction of the initial problem mode failure intensity

DOT&E TEMP Guide 3.0

- **Provides guidance on incorporation of the Program's Reliability Growth Strategy in the TEMP**

- **Requires that the TEMP include an overview of the reliability program and testing needed to assess/monitor reliability growth, including design for reliability T&E activities.**

- **Requires a brief description of key engineering activities supporting the reliability growth program:**
  - Reliability allocations to components and subsystems,
  - Reliability block diagrams (or system architectures for software intensive systems) and predictions
  - Failure definitions and scoring criteria (FDSC)
  - Failure mode, effects and criticality analysis (FMECA)
  - System environmental loads and expected use profiles
  - Dedicated test events for reliability such as accelerated life testing, and maintainability and built-in test demonstrations
  - Reliability growth testing at the system and subsystem level
  - Failure reporting analysis and corrective action system (FRACAS) maintained through design, development, production, and sustainment.

DOT&E TEMP Guide 3.0 (cont.)

- **The reliability growth program described in the TEMP should contain the following**
  - Initial estimates of system reliability and a description of how this estimates were arrived at
  - Reliability growth planning curves (RGPC) illustrating the reliability growth strategy, and including justification for assumed model parameters (e.g. fix effectiveness factors, management strategy)
  - Estimates with justification for the amount of testing required to surface failure modes and grow reliability
  - Sources of sufficient funding and planned periods of time to implement corrective actions and test events to confirm effectiveness of those actions
  - Methods for tracking failure data (by failure mode) on a reliability growth tracking curve (RGTC) throughout the test program to support analysis of trends and
  - changes to reliability metrics
  - Confirmation that the Failure Definition Scoring Criteria (FDSC) on which the RGPC is based is the same FDSC that will be used to generate the RGTC
  - Entrance and exit criteria for each phase of testing Operating characteristic (OC) curves that illustrate allowable test risks (consumer's and producer's risks) for assessing the progress against the reliability requirement. The risks should be related to the reliability growth goal.

Reliability Growth Projection
(AMSAA Crow Projection Model)

Projected Failure Intensity	=	Failure intensity at end of test	−	Reduction in failure intensity due to corrective actions	+ d	Failure intensity for unseen BD-modes you intend to fix in future
λ_p		λ_s		$-\lambda_{bd} + \left(\sum_{j=1}^{N}(1-d_j)\,\lambda_j\right)$		$dh(t)$

where:

- λ_s is the failure intensity at the end of the test of length T (# failures / T)
- N is the number of unique BD-modes seen by T
- d_j is the individual failure mode fix effectiveness for BD-mode j
- d is the average fix effectiveness for all BD-modes
- λ_j is the failure intensity for BD-mode j
- λ_{bd} is the failure intensity for BD-modes being corrected
- $h(t)$ is the failure intensity of unseen BD modes at time t (also known as discovery rate):

$$h(t) = \hat{\lambda}\,\hat{\beta}\,t^{\hat{\beta}-1} \quad \text{where} \quad \hat{\beta} = \frac{N-1}{\sum_{i=1}^{N}\ln\left(\frac{T}{X_i}\right)} \qquad \hat{\lambda} = \frac{N}{T^{\hat{\beta}}}$$

  - $\hat{\lambda}$ is the AMSAA-Crow model scale parameter
  - $\hat{\beta}$ is the AMSAA-Crow model shape parameter
  - X_i is the i-th successive failure time (considering first occurrence times for each BD mode)

Reliability Growth Projection (cont.)
(AMSAA Crow Projection Model)

Suppose a system demonstrated the following failures in a 500-hour test:

Test Time (hours)	500
Total Failures	42
Number of A-mode Failures	8
Number of BD-mode Failures	34
Number of Unique BD-modes	16

BD Mode	BD Mode First Occurrence times (X_i)	No. of failures (n_i)	Estimated FEF (d_j)	Adjusted Failure Intensity for each BD Mode ($(1-d_j)\,\lambda_j$)
1	14	3	0.77	0.0014
2	27.2	4	0.63	0.0030
3	57	1	0.65	0.0007
4	63	1	0.78	0.0004
5	99	2	0.9	0.0004
6	125	2	0.5	0.0020
7	159	1	0.85	0.0003
8	176	1	0.85	0.0003
9	180	4	0.72	0.0022
10	181.5	5	0.72	0.0028
11	275	3	0.7	0.0018
12	295	2	0.63	0.0015
13	356	1	0.75	0.0005
14	379.4	2	0.8	0.0008
15	450	1	0.65	0.0007
16	485	1	0.5	0.0010

Calculations

Mean Time Between Failure	11.90
System Failure Intensity (λ_s)	0.084
BD-mode Failure Intensity (λ_{bd})	0.068
A-mode failure intensity	0.016
Average FEF	0.713

Adjusted BD-mode Failure Intensity for all BD-modes

$$\sum_{j=1}^{N}(1-d_j)\,\lambda_j$$

0.020

AMSAA Crow Shape Parameter

$$\hat{\beta} = \frac{N-1}{\sum_{i=1}^{N}\ln\left(\frac{T}{X_i}\right)}$$

0.756

AMSAA Crow Scale Parameter

$$\hat{\lambda} = \frac{N}{T^{\hat{\beta}}}$$

0.146

Failure intensity of unseen BD-modes at T=500 hours

$$h(t) = \hat{\lambda}\hat{\beta}t^{\hat{\beta}-1}$$

0.024

Reliability Projection Using Correct and Incorrect Approaches

1) Correct Projection Method: Consider FEF and unseen BD-mode failure rate

   $\lambda_p =$ 0.084 - 0.068 + 0.020 + 0.024 (0.713) = 0.0531

   MTBF = 18.8 hours

2) Incorrect Projection Method 1: Ignore FEF and unseen BD-mode failure rate

   $\lambda_p =$ 0.084 - 0.068 = 0.016

   MTBF = 62.5 hours

3) Incorrect Projection Method 2: Account for FEF but ignore unseen BD-mode failure rate

   $\lambda_p =$ 0.084 - 0.068 + 0.020 = 0.036

   MTBF = 27.8 hours

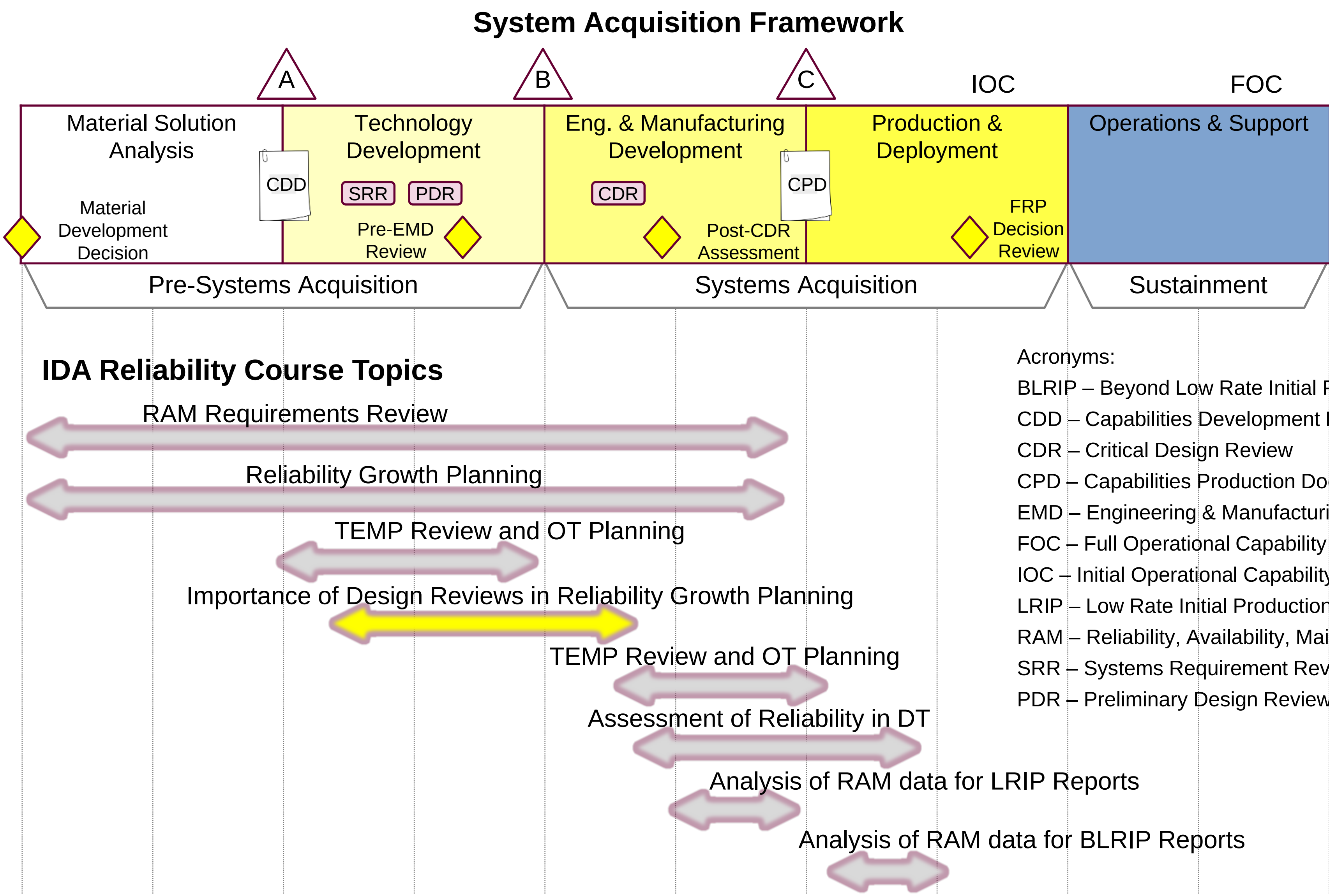
Institute for Defense Analyses
4850 Mark Center Drive • Alexandria, Virginia 22311-1882

Importance of Design Reviews in the Reliability Growth Planning Process

Jonathan L. Bell

2 June 2016

Timeline



Acronyms:

BLRIP – Beyond Low Rate Initial Production

CDD – Capabilities Development Document

CDR – Critical Design Review

CPD – Capabilities Production Document

EMD – Engineering & Manufacturing Development

FOC – Full Operational Capability

IOC – Initial Operational Capability

LRIP – Low Rate Initial Production

RAM – Reliability, Availability, Maintainability

SRR – Systems Requirement Review

PDR – Preliminary Design Review

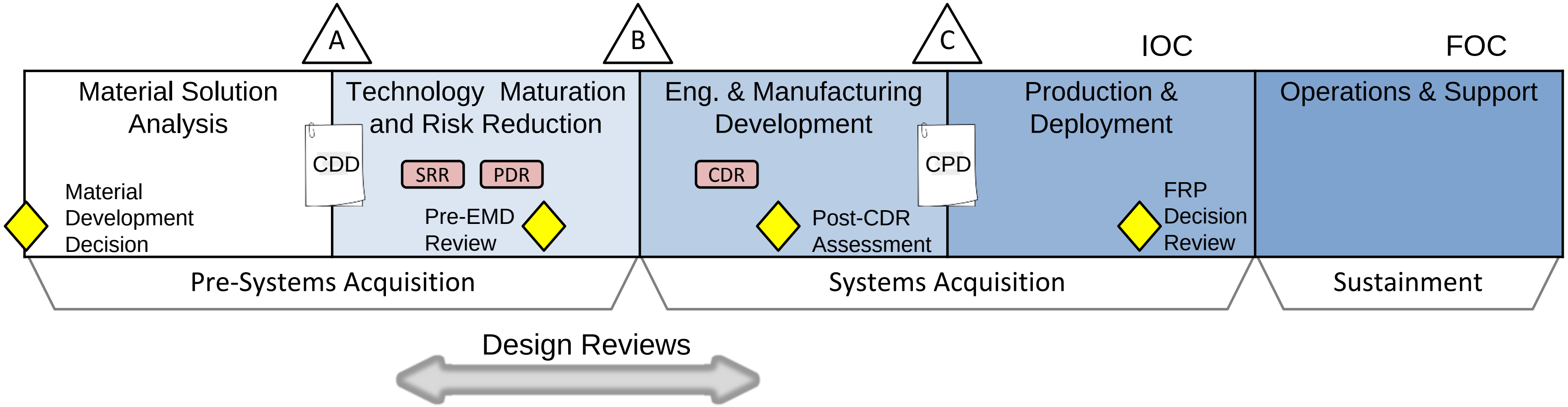
Purpose of Briefing

- Highlight the importance of design reviews in the reliability growth planning process.
- Identify the relevant questions to consider during design review activities.
- Provide programmatic examples of this process.

Outline

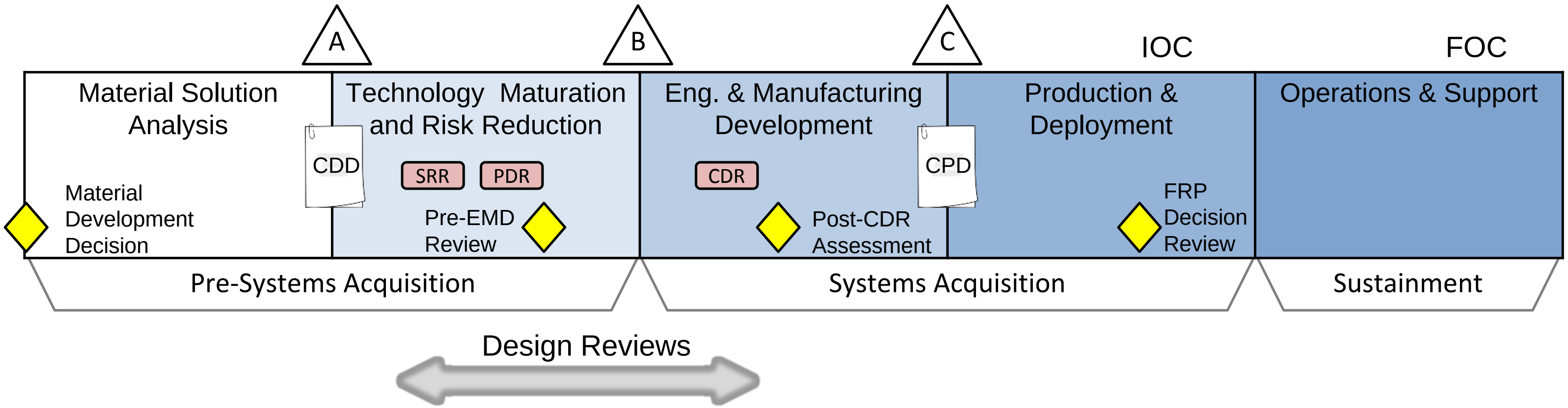
- **Overview of Design Reviews**
- **Key Reliability Growth Planning Considerations During Design Reviews**
- **Programmatic Examples**

Overview of Design Reviews



- **A detailed understanding of the system's design and the developer's system engineering process is critical to building a credible reliability growth strategy.**
- **Per DOD 5000.02, dated 7 January 2015, "any program that is not initiated at Milestone C will include the following design reviews":**
- **Preliminary Design Review (PDR):**
  - Assesses the maturity of the preliminary design supported by the results of requirements trades, prototyping, and critical technology demonstrations. The PDR will establish the allocated baseline and confirm that the system under review is ready to proceed into detailed design.
- **Critical Design Review (CDR)**
  - Assesses design maturity, design build-to or code-to documentation, and remaining risks and establishes the initial product baseline. Used as the decision point that the system design is ready to begin developmental prototype hardware fabrication or software coding with acceptable risk.

Overview of Design Reviews (cont.)



- **Per DOD 5000.02, dated 7 January 15, the Program Manager will formulate a comprehensive Reliability and Maintainability program to ensure reliability and maintainability requirements are achieved; the program will consist of engineering activities including for example":**
  - R&M allocations
  - Block diagrams and predictions
  - Failure definitions and scoring criteria
  - Failure mode, effects and criticality analysis
  - Maintainability and built-in test demonstrations
  - Reliability testing at the system /subsystem level
  - Failure reporting, analysis, and corrective action system maintained through design, development, production, and sustainment

- **Reliability growth planning is an integral part of the systems engineering process.**
- **In addition to design reviews, contract deliverables, developed early in a program, might also provide documentation on the system design and the extent that the contractor had included reliability in the systems engineering process.**

Key Reliability Growth Planning Questions to Address

- **Are the reliability requirement(s) understood by the developer?**
  - Are reliability goal(s) included in contractual documents?
  - Is the reliability growth goal linked to the user's reliability requirement, if applicable?
  - Is the developer aware of interim reliability goals such as entrance/exit criteria for various test phases, if applicable?
  - Has the failure definition and/or scoring criteria been communicated to the developer? For software, has the defect prioritization been defined?
  - Does the developer have reliability test data that can be assessed to verify compliance with the Government's scoring process?

- **Are reliability predictions credible?**
  - Does the developer have an estimate for the initial reliability of the system/subsystems? If so, is the estimate consistent with the reliability growth planning curve?
  - Are predictions supported by test data that are based on use of the system over its representative mission profile and scoring of failures in accordance with approved failure definition and/or scoring criteria?
  - Was testing and data collection performed by a government test site?
  - Does developer have a reliability block diagram?
  - Were reliability predictions based on MIL-STD-217 or is progeny (common on space programs)?
  - Were reliability predictions based on a physics of failure model?
  - Did the contractor implement a Design for Reliability (DfR) process?
  - Does the developer have a history of producing reliability hardware/software?

Key Reliability Growth Planning Questions to Address (cont.)

- **Is the developer's Management Strategy (MS)* credible?**
  - Is there adequate funding and time to discover failure modes and develop, implement, and verify corrective actions.
  - How mature is the design/software code? Is the design a new build? Does it incorporate Commercial Off-the-Shelf (COTS), Government Off-the-Shelf (GOTS), or Government Furnished Equipment (GFE)?
  - Will the program address failures due to COTS/GOTS/GFE or borrowed software code? If not, were these subsystems/components/code included as part of the A-mode failure intensity.
  - Is there representative field failure data on the subsystems/components/software? If so, has this data been appropriately scored in accordance with the the failure definition and/or scoring criteria? Was this information used to develop an estimate for MS?

- **How mature is the system that will enter testing?**
  - When will a functional prototype or fully function software code be available?
  - Has the developer conducted testing of the system on their own?
  - Does the program anticipate major design/manufacturing changes or software drops after MS C?

- **Is the developer required to conduct break-in or shakedown testing?**
  - If so, are there specific criteria that should be met?
  - What is the mitigation plan if the developer fails to meet break-in or shakedown criteria?

Management Strategy* $$MS = \frac{\lambda_B}{\lambda_A + \lambda_B}$$

λ_B = initial B-mode failure intensity

λ_A = initial A-mode failure intensity

Programmatic Examples

- Design reviews provide information that is essential to reliability growth planning
- This section provides programmatic examples, including the following:
  - Reliability Planning for System Upgrades (x2)
  - Importance of Understanding Scoring Criteria
  - It Doesn’t Matter What We Did Before
  - Growth to infinity

AH-64E Apache

OH-58F Kiowa Warrior

Joint Light Tactical Vehicle

F-15 Radar Modernization Program

"Reliability Planning For System Upgrades"

- **OH-58F Kiowa Warrior**
  - During the System Requirement Review and subsequent Preliminary Design Review, DOT&E learned that most of OH-58F parts were not new; they came from the legacy OH-58D aircraft
  - Program office stated they would not implement corrective actions for any of the legacy components
  - Initial program growth curve had a 0.95 Management Strategy (MS), which is typical of a new start program.
  - DOT&E obtained detailed failure mode data from the program office on legacy and new system components.
  - Analysis of the failure mode data indicated that a 0.5 MS was more realistic.

$$MS = \frac{\lambda_B}{\lambda_A + \lambda_B}$$

λ_B = initial B-mode failure intensity

λ_A = initial A-mode failure intensity

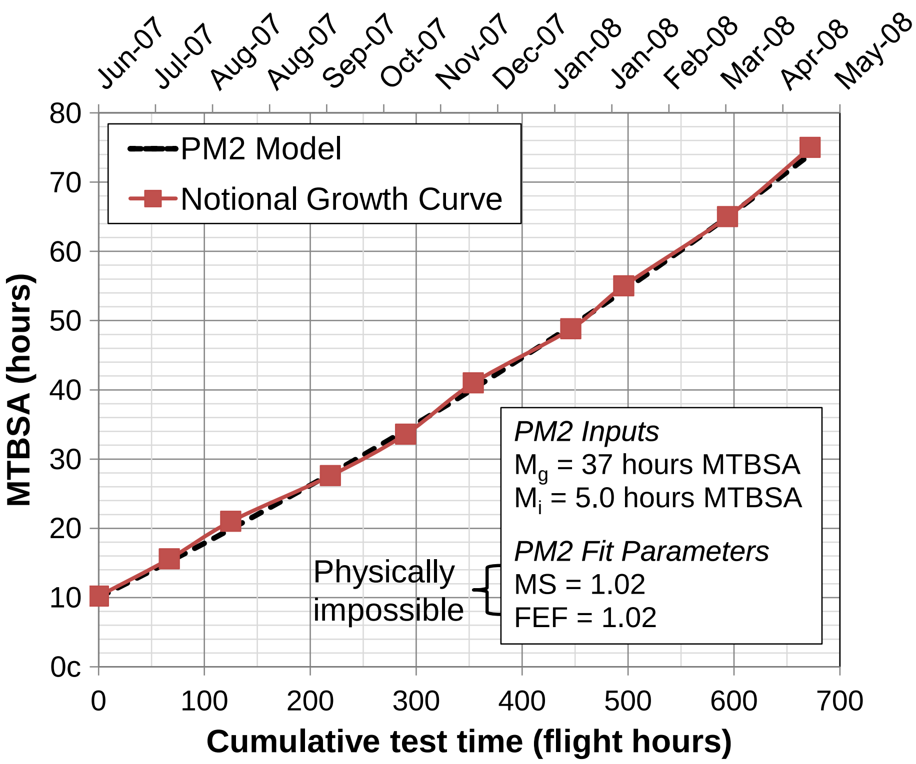
Ensure estimates of growth and management strategy are realistic. They should accurately quantify what the program intends to fix.

"Reliability Planning For System Upgrades"

- **F-15E Radar Modernization Program (RMP)**
  - RMP initially had a hardware reliability requirement only
  - For AESA radars, software accounts for the majority of failures
  - Program established Mean Time Between Software Anomalies (MTBSA) requirement
  - RMP software code maturity

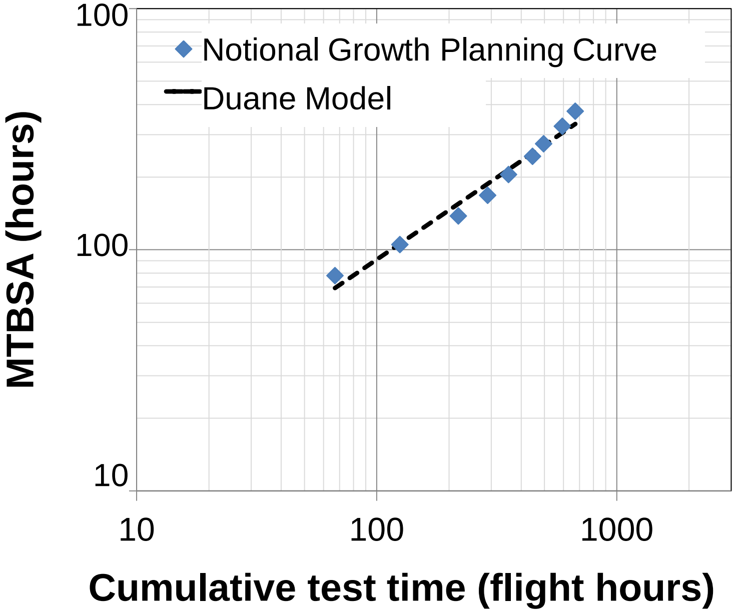
- **DOT&E and IDA assessed the programs stability growth curve as overly aggressive**

PM2 Model Fit to Notional Contractor Curve



Acronyms:
FEF – Fix Effectiveness Factor
MS – Management Strategy
Mi – Initial Reliability
M_g – Reliability Growth Goal
PM2 – Planning Model based on Projection Methodology

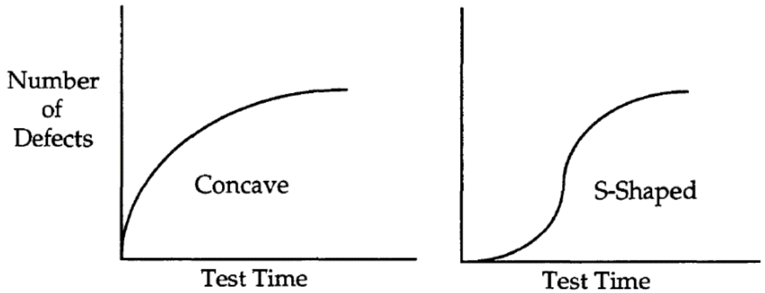
“Reliability Planning For System Upgrades (cont)”



- **Comparison of notional curve to Duane model suggests that growth curve projections are aggressive**
- **Fitted growth rate parameter (α) ~ 0.70**

Military Standard 189C:

- Historical mean/median for α is 0.34/0.32
- Historical range for α is 0.23 - 0.53
- An α of 0.70 is unrealistically aggressive, particularly for a program that is incorporating mostly mature technology

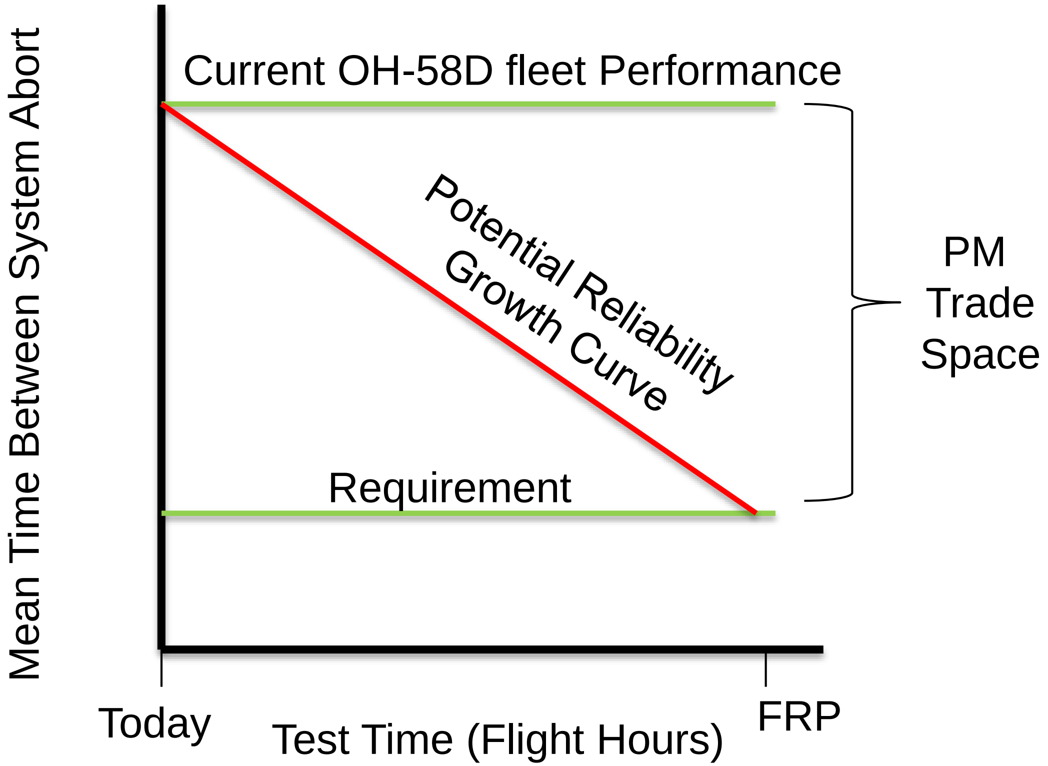


Ensure reliability growth estimates are realistic. They should accurately quantify what the program intends to fix.

MTBSA – Mean Time Between Software Anomalies

“Understanding Scoring Criteria”

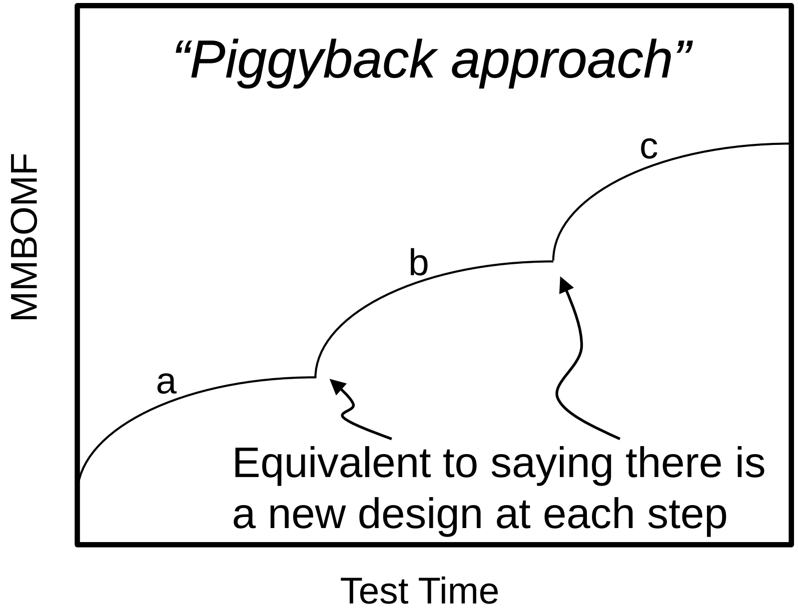
- **OH-58F Kiowa Warrior**
  - Reliability requirement based on 1990s document
  - OH-58D had multiple upgrades and reliability improvements since 1990
  - Combat reliability estimates were much higher than the requirement
  - Rescored combat data with Failure Definition Scoring Criteria (FDSC) to obtain a more accurate reliability estimate
    - Estimated reliability of current system exceeded requirement.



Ensure initial reliability estimate reflects the reliability of the current system considering all engineering changes made over the years.

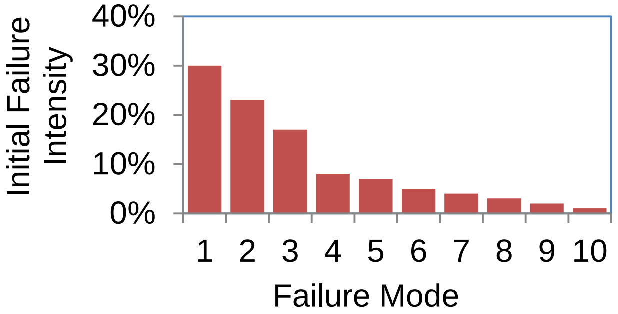
"It Doesn't Matter What We Did Before"

- **Joint Light Tactical Vehicle (JLTV)**
  - The early JLTV TEMP included three growth curves projecting growth out to the objective reliability requirement for Mean Miles Between Operational Mission Failure (MMBOMF):



Problems with this approach

- Subsequent steps overestimate the growth that can be achieved since the bulk of high rate failure modes were already addressed in the first step
- Steps "b" and "c" essentially assume system redesigns



Make sure the reliability growth curves are based on realistic assumptions.

"Growth to Infinity"

- **F-15E Radar Modernization Program (RMP)**
  - Had a Mean Time Between Critical Failure (MTBCF) requirement at Full Operational Capability (FOC)
  - Used Duane model reliability growth planning curve
- **Duane Model is more appropriate for tracking/analysis vice reliability growth planning because it has the following limitations:**
  - Permits growth to infinity as $t \rightarrow \infty$
  - Growth potential not considered
  - Converges to zero as $t \rightarrow 0$
  - 100% fix effectiveness
  - Growth not linked to engineering or management

Ensure reliability growth curve is based on realistic assumptions that are tied to engineering, program management, and the test plan.

“Mission Aborts in DT”

- **Programs typically build reliability growth strategy/curves for mission failure or mission abort requirement**
- **Mission aborts occur less frequently than Essential Function Failures (EFFs) or Essential Maintenance Actions (EMAs)**
- **Growth strategies based on EMAs produce a more credible and less resource-intensive reliability growth strategy by:**
  - Incorporating a larger share of the failure modes
  - Addressing problems before they turn into mission aborts
  - Improving the ability to assess and track reliability growth
  - Increasing the statistical power and confidence to evaluate reliability in testing
  - Enabling more reasonable reliability growth goals
  - Reducing subjectivity that can creep into the reliability scoring process
- **AH-64E decided to focus growth strategy on Mean Time Between EMAs as well as Mean time between Mission Aborts**

Takeaway Points

- **Get involved early in developing reasonable estimates for growth parameters**
  - Participate in design reviews to understand proposed design.
    - The design for a system upgrade might have changed many times over the years (e.g., OH-58F)
  - Work with Reliability Integrated Product Team to ensure growth parameters are tied to engineering, contracting documentation, program management, and the test plan
- **Discuss requirements: KPPs are not always the best for reliability growth planning curves**
  - Fight inadequate requirements (e.g., F-15 Radar Modernization Program (RMP) Full Operational Capability reliability requirement)
  - In the absence of adequate requirements, compare to legacy performance in testing (e.g., OH-58F Kiowa Warrior)
  - Push for reliability growth planning curves based on EMAs/EFFs
- **Build a realistic reliability growth plan that is based on systems engineering**
  - Ensure it considers the reliability growth potential and does not permit infinite growth (e.g., Duane model)
  - Ensure it represents the specific failure modes the program intends to fix. It should consider all A-modes, particularly for non new-start systems (e.g., OH-58F, F-15E RMP radar software)
  - Confirm that it is supported with a Failure Reporting and Corrective Action System and Failure Review Board
  - Update model inputs once test results are available
  - Ensure design margins are adequate

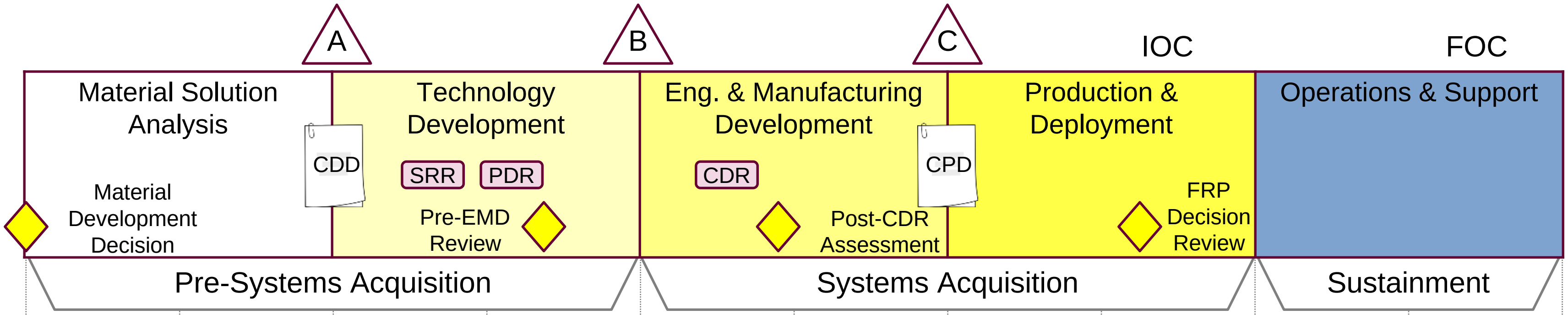
Institute for Defense Analyses
4850 Mark Center Drive • Alexandria, Virginia 22311-1882

TEMP Review and OT Planning

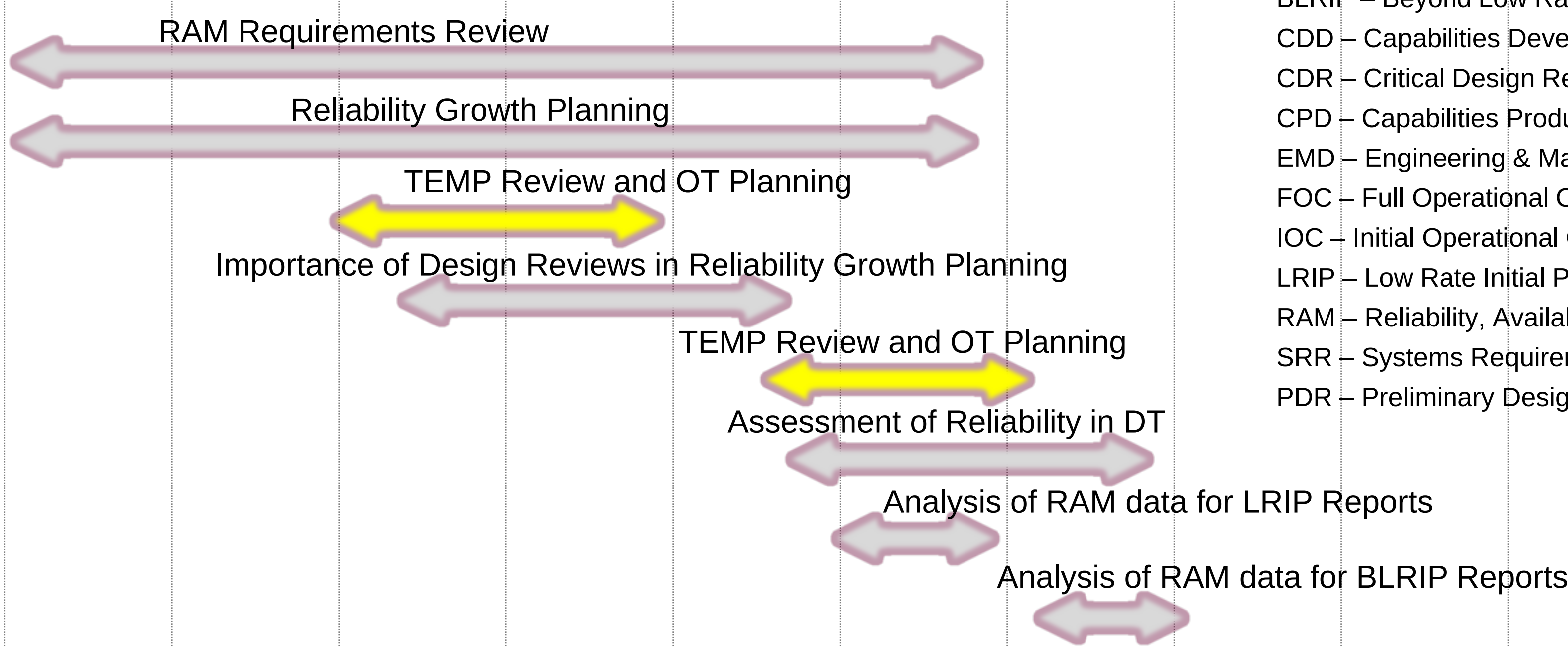
Rebecca Dickinson
2 June 2016

Timeline

System Acquisition Framework



IDA Reliability Course Topics



Acronyms:

BLRIP – Beyond Low Rate Initial Production

CDD – Capabilities Development Document

CDR – Critical Design Review

CPD – Capabilities Production Document

EMD – Engineering & Manufacturing Development

FOC – Full Operational Capability

IOC – Initial Operational Capability

LRIP – Low Rate Initial Production

RAM – Reliability, Availability, Maintainability

SRR – Systems Requirement Review

PDR – Preliminary Design Review

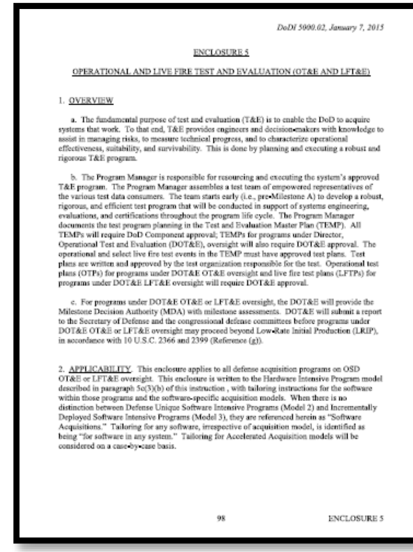
Purpose of Briefing

- How should programs document their reliability growth plan in the TEMP?
- What criteria that should be considered during the review process?
- How do we assess the adequacy of an OT ?

Reliability is the chief enabler of operational suitability, and failure to achieve reliability requirements typically results in a system being assessed "not suitable"; consequently, its independent evaluation is pivotal to OT&E.
Independent Operational test and Evaluation (OT&E) Suitability Assessments – October 05 2012 DOT&E Memo

OT&E for Reliability

- The TEMP must include a plan (typically via a working link to the Systems Engineering Plan) to allocate reliability requirements down to components and sub-components.
- Beginning at Milestone B, the TEMP must include Test & Evaluation (T&E) for reliability growth and reliability growth curves (RGCs) for the whole system and the reliability of critical systems, sub-systems, components, and sub-components.
- RGCs must display planned initial reliability, the allocated reliability requirement, a curve showing reliability that is expected during each reliability test event, and points marking reliability test results to date.
- Beginning at Milestone B, the TEMP must include a working link to the failure mode, effects and criticality analysis (FMECA)
- Updated TEMPs at Milestone C must include updated RGCs that reflect test results to date, any updates to the planned T&E for reliability growth, and a working link to the updated FMECA.



DoDI 5000.02, January 7, 2015

ENCLOSURE 5

OPERATIONAL AND LIVE FIRE TEST AND EVALUATION (OT&E AND LFT&E)

1. OVERVIEW

a. The fundamental purpose of test and evaluation (T&E) is to enable the DoD to acquire systems that work. To that end, T&E provides engineers and decision-makers with knowledge to assist in managing risks, to measure technical progress, and to characterize operational effectiveness, suitability, and survivability. This is done by planning and executing a robust and rigorous T&E program.

b. The Program Manager is responsible for resourcing and executing the system's approved T&E program. The Program Manager assembles a test team of empowered representatives of the various test data consumers. The team starts early (i.e., pre-Milestone A) to develop a robust, rigorous, and efficient test program that will be conducted in support of systems engineering, evaluations, and certifications throughout the program life cycle. The Program Manager documents the test program planning in the Test and Evaluation Master Plan (TEMP). All TEMPs will require DoD Component approval; TEMPs for programs under Director, Operational Test and Evaluation (DOT&E), oversight will also require DOT&E approval. The operational and select live fire test events in the TEMP must have approved test plans. Test plans are written and approved by the test organization responsible for the test. Operational test plans (OTPs) for programs under DOT&E OT&E oversight and live fire test plans (LFTPs) for programs under DOT&E LFT&E oversight will require DOT&E approval.

c. For programs under DOT&E OT&E or LFT&E oversight, the DOT&E will provide the Milestone Decision Authority (MDA) with milestone assessments. DOT&E will submit a report to the Secretary of Defense and the congressional defense committees before programs under DOT&E OT&E or LFT&E oversight may proceed beyond Low-Rate Initial Production (LRIP), in accordance with 10 U.S.C. 2366 and 2399 (Reference (g)).

2. APPLICABILITY. This enclosure applies to all defense acquisition programs on OSD OT&E or LFT&E oversight. This enclosure is written to the Hardware Intensive Program model described in paragraph 5c(3)(b) of this instruction , with tailoring instructions for the software within those programs and the software-specific acquisition models. When there is no distinction between Defense Unique Software Intensive Programs (Model 2) and Incrementally Deployed Software Intensive Programs (Model 3), they are referenced herein as "Software Acquisitions." Tailoring for any software, irrespective of acquisition model, is identified as being "for software in any system." Tailoring for Accelerated Acquisition models will be considered on a case-by-case basis.

98 ENCLOSURE 5

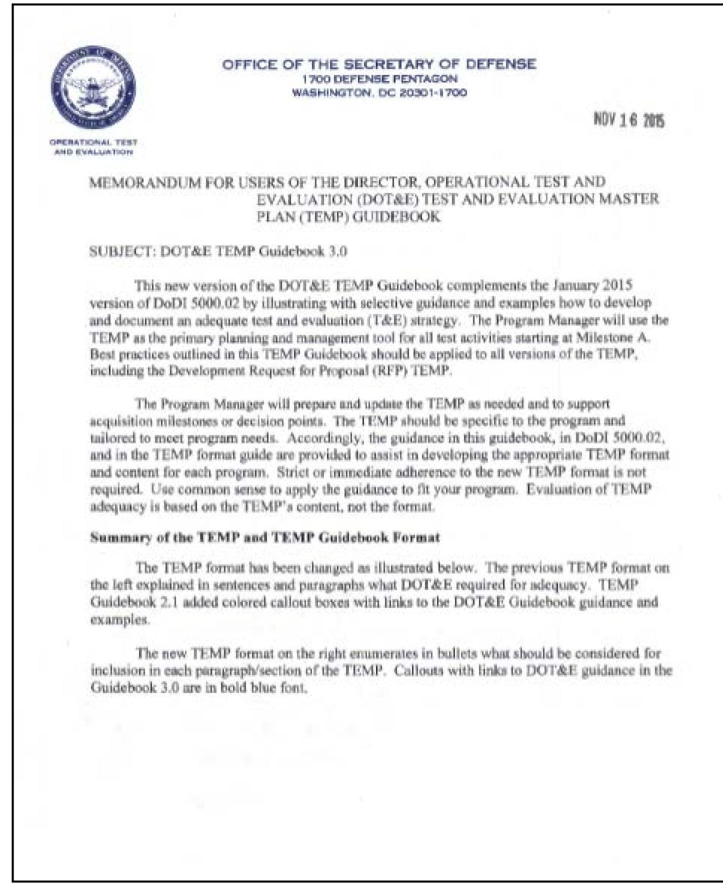
Reliability and Maintainability Policy: DoDI 5000.02

Reliability Growth in the TEMP

The DOT&E TEMP Guidebook Version 3.0

– Provides guidance on the incorporation of the program's reliability growth plan in the TEMP

– Requires that the TEMP include an overview of the reliability program and testing needed to assess/monitor reliability growth, including design for reliability T&E activities



OPERATIONAL TEST AND EVALUATION

OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

NOV 16 2015

MEMORANDUM FOR USERS OF THE DIRECTOR, OPERATIONAL TEST AND EVALUATION (DOT&E) TEST AND EVALUATION MASTER PLAN (TEMP) GUIDEBOOK

SUBJECT: DOT&E TEMP Guidebook 3.0

This new version of the DOT&E TEMP Guidebook complements the January 2015 version of DoDI 5000.02 by illustrating with selective guidance and examples how to develop and document an adequate test and evaluation (T&E) strategy. The Program Manager will use the TEMP as the primary planning and management tool for all test activities starting at Milestone A. Best practices outlined in this TEMP Guidebook should be applied to all versions of the TEMP, including the Development Request for Proposal (RFP) TEMP.

The Program Manager will prepare and update the TEMP as needed and to support acquisition milestones or decision points. The TEMP should be specific to the program and tailored to meet program needs. Accordingly, the guidance in this guidebook, in DoDI 5000.02, and in the TEMP format guide are provided to assist in developing the appropriate TEMP format and content for each program. Strict or immediate adherence to the new TEMP format is not required. Use common sense to apply the guidance to fit your program. Evaluation of TEMP adequacy is based on the TEMP's content, not the format.

Summary of the TEMP and TEMP Guidebook Format

The TEMP format has been changed as illustrated below. The previous TEMP format on the left explained in sentences and paragraphs what DOT&E required for adequacy. TEMP Guidebook 2.1 added colored callout boxes with links to the DOT&E Guidebook guidance and examples.

The new TEMP format on the right enumerates in bullets what should be considered for inclusion in each paragraph/section of the TEMP. Callouts with links to DOT&E guidance in the Guidebook 3.0 are in bold blue font.

Key Engineering Activities

- **The TEMP requires a brief description of key engineering activities that support the reliability growth program:**
  - Reliability allocations to components and subsystems
  - Reliability block diagrams (or system architectures for software intensive systems) and predictions
  - Failure definitions and scoring criteria (FDSC)
  - Failure mode, effects and criticality analysis (FMECA)
  - Systems environmental loads and expected use profiles
  - Dedicated test events for reliability such as accelerated life testing, and maintainability and built-in test demonstrations
  - Reliability growth testing at the system and subsystem level
  - A failure reporting analysis and corrective action system (FRACAS) maintained through design, development, production, and sustainment

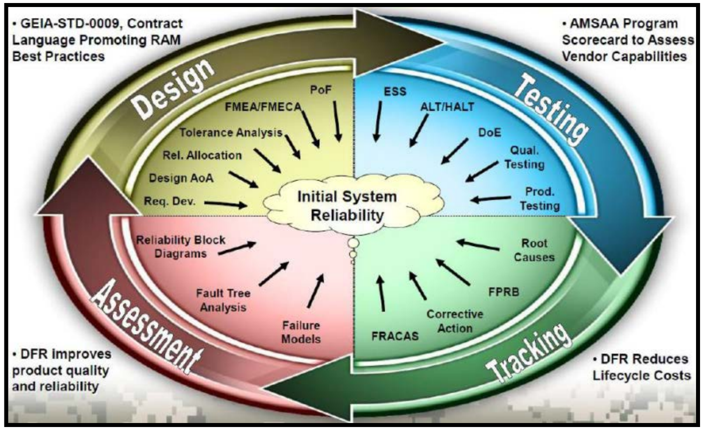
- **The key engineering activities should be discussed in much more detail in the appropriate supporting references. References, such as the System Engineering Plan or the Reliability Program Plan, should be provided in the TEMP**

Reliability Growth Program

- **The TEMP should contain the following information with respect to the reliability growth program:**
  - Initial estimates of system reliability and how estimates were determined
  - Reliability growth planning curves (RGPC) illustrating the growth strategy, and justification for assumed model parameters (fix effectiveness factors, management strategy, corrective actions)
  - Estimates with justification for the amount of testing required to surface failure modes and grow reliability
  - Methods for tracking failure data (by failure mode) on a reliability growth tracking curve (RGTC) throughout the test program to support analysis of trends and changes to reliability metrics
  - Confirmation that the FDSC on which the RGPC is based is the same FDSC that will be used to generate the RGTC
  - Entrance and exit criteria for each phase of testing
  - Operating characteristic curves that illustrate allowable test risks (consumer's and producer's risks) for assessing the progress against the reliability requirement. The risks should be related to the reliability growth goal.

Reliability growth curves are excellent planning tools, but programs will not achieve their reliability goals if they treat reliability growth as a "paper policy." Good reliability planning must be backed up by sound implementation and enforcement.
(DOT&E FY 2014 Annual Report)

Reliability: Design and Growth



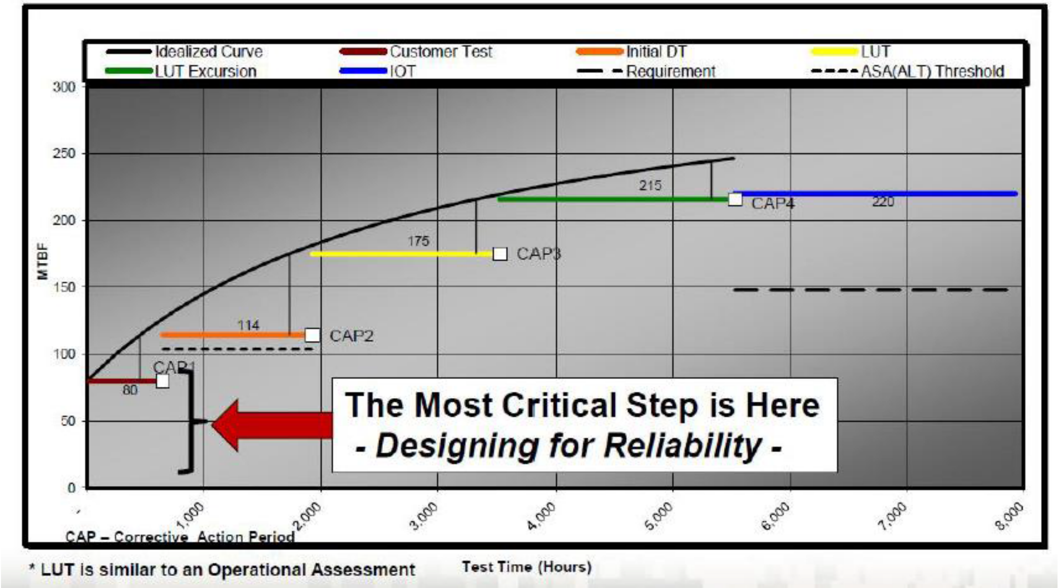
- **Reliability Growth:** The positive improvement in a reliability parameter over a period of time due to changes in the product design or manufacturing process (MIL-HDBK 189C, 2011).

Testing alone will not improve reliability – only corrective actions can improve reliability!

- **Reliability Growth Management:** The systematic planning for reliability achievement as a function of time and other resources, and controlling the ongoing rate of achievement by reallocation of resources based on comparisons between planned and achieved reliability values (MIL-HDBK 189C, 2011)

Design for Reliability, then Reliability Growth!

Reliability Growth Curve – Projection Model 2



"Get on the curve and stay on the curve" - AMSAA

Graphic sources: AMSAA Reliability Growth Short Course

Reliability Growth Tracking

- **Does the TEMP describe how reliability will be tracked across the developmental life cycle?**
- **Why is tracking important?**
  - Determine if growth is occurring and to what degree.
  - Estimate the demonstrated reliability based on test data
  - Compare the demonstrated reliability to the requirements
- **How do we track reliability?**
  - The most common methods of growth tracking are scoring and assessment conferences, measures to determine if reliability is increasing in time, tracking models, and FRACAS.
- **Defense Acquisition Executive Summary requires Major Defense Acquisition Programs with a documented reliability growth curve in the SEP or TEMP to report reliability data on a quarterly basis.**
- **Systems not meeting entrance and exit criteria should revise the reliability growth strategy to reflect current system reliability**

Reliability should be measured, monitored, and reported throughout the acquisition process.

Reliability Entrance and Exit Criteria

Systems not meeting entrance and exit criteria should revise the reliability growth strategy to reflect current system reliability

What are the intermediate goals or entrance and exit goals?
Is there enough time planned in each test phase to surface failure modes?
Are there planned CAPs at the end of phase?
Are the CAPs of reasonable length?
If a requirement is not met – will a CAP be initiated?

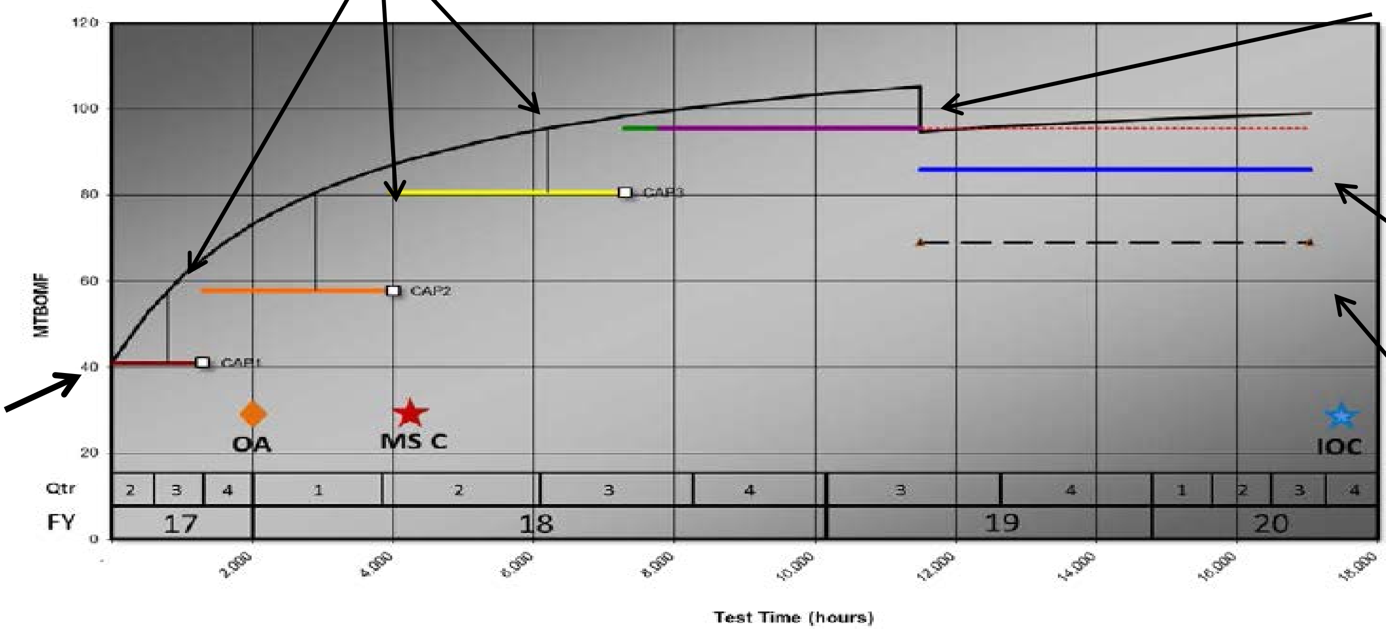
Is the DT MTBOMF on contract?

DT Goal = 95 MTBF

IOT&E Goal = 86 MTBF

Requirement = 69 MTBF

Will we even start on the curve?



Are reliability entrance criteria specified for the OA?
Are they consistent with the curve?

TEMP Example

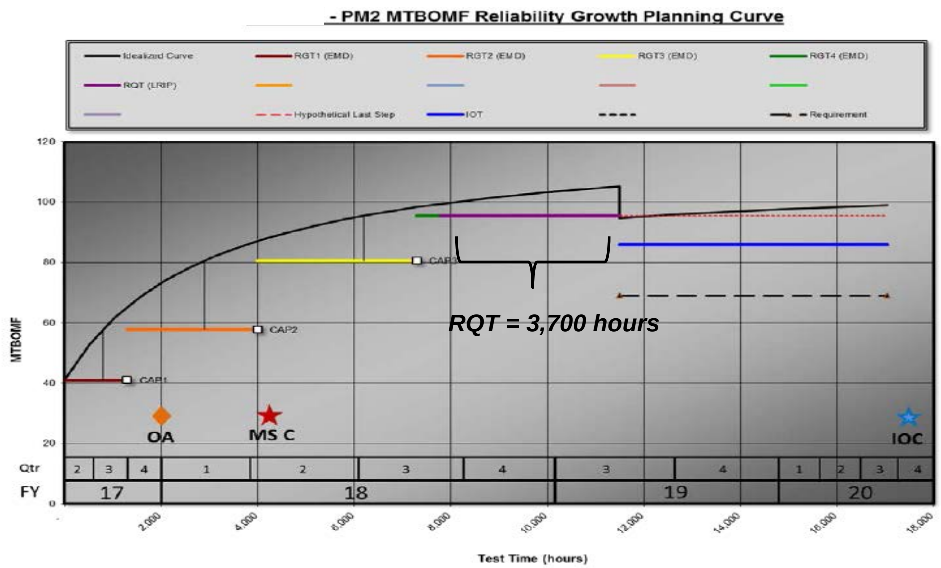
- **System Engineering Strategy**
  - Engineering Activities: R&M allocations; block diagrams and predictions; FMECA; FDSC
- **Comprehensive Growth Plan Overview outlined in TEMP**
  - Described for both MTBOMF and MTBF
    - » Provide adequate justification for initial system level reliability, Management Strategy, FEF, CAPs, etc.
- **Pre-IOT&E Reliability Qualification Test**
  - Program Office wants to evaluate system reliability prior to IOT&E.
    - » Expected 69 hour MTBOMF will be demonstrated with 80% confidence and have a 70% probability of acceptance during RQT
- **IOT&E is long enough**
  - If the vehicle meets it DT reliability growth goal of 95 hours the IOT&E will be long enough to demonstrate the 69-hour MTBFOM requirement with 80% confidence and 84% power (assuming a 10% degradation from DT to OT).
- **Growth Goal is on Contract!**
  - The reliability growth goal was included in the program's Request for Proposals!

Planning Inputs	
Requirement MTBOMF (M_R)	69.0 Hours
Initial MTBOMF (M_I)	41.0 Hours
Management Strategy (MS)	0.95
Average FEF (μ_d)	0.70
Confidence Level for IOT LCB	0.80
Prob. Of Accept. In IOT using LCB	0.84
IOT Inputs	
IOT Phase Time	5,544 Hours
DT to IOT Degradation Factor	0.10
Outputs	
Prob. Of Accept. In IOT using Pt. Est.	0.97
Goal MTBOMF in IOT (M_R^+)	86 Hours
Goal MTBOMF in IOT (M_G)	95 Hours
Growth Potential (M_{GP})	122 Hours
M_G/M_{GP} Ratio	0.78

Reliability Growth Curve Assumptions

Developmental Test - Schedule Summary						
dule ase	Test Phase Name	Mission Time in Test Phase	Cumulative Test Time	CAP at End of Phase?	Corrective Action Lag Time for Individual CAP	Cumulative Time at CAP minus Lag
le	RGT1 (EMD)	1,300	1,300	Yes	500	800
le	RGT2 (EMD)	2,700	4,000	Yes	1,100	2,900
le	RGT3 (EMD)	3,300	7,300	Yes	1,100	6,200
le	RGT4 (EMD)	500	7,800	No		6,200
le	RQT (LRIP)	3,700	11,500	No		6,200
le						
TP 7 Schedule						
TP 8 Schedule						
TP 9 Schedule						
TP 10 Schedule						

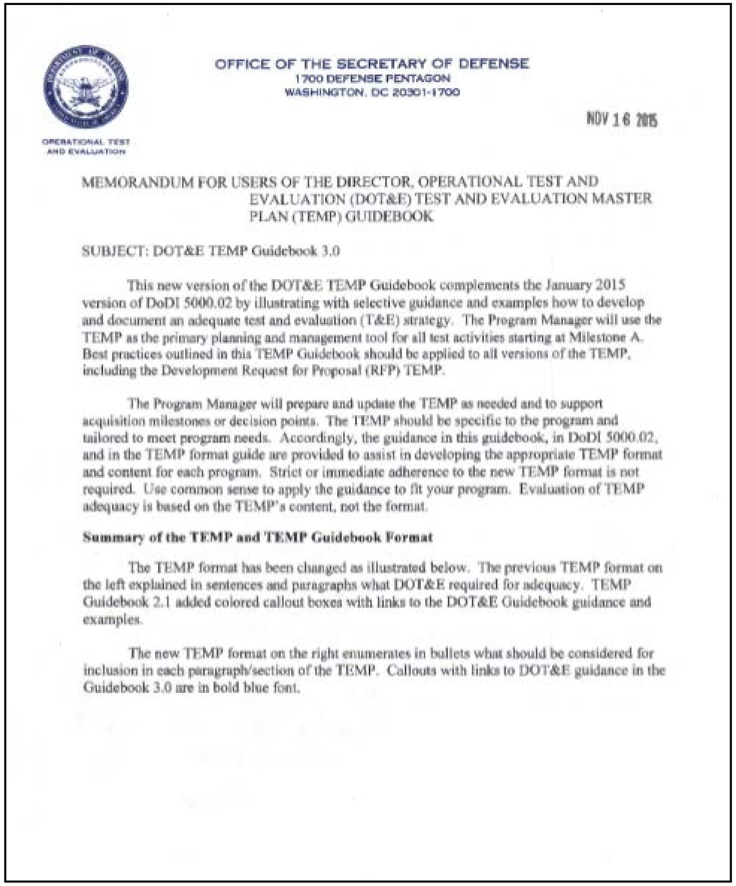
Projected Miles Supporting Reliability Growth



Reliability Test Planning

The DOT&E TEMP Guidebook Version 3.0

- Provides guidance on the Reliability Test Planning
  - » Reliability Requirements
  - » Planning an Adequate Test
  - » Incorporating Additional Information



OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST AND EVALUATION

NOV 16 2015

MEMORANDUM FOR USERS OF THE DIRECTOR, OPERATIONAL TEST AND EVALUATION (DOT&E) TEST AND EVALUATION MASTER PLAN (TEMP) GUIDEBOOK

SUBJECT: DOT&E TEMP Guidebook 3.0

This new version of the DOT&E TEMP Guidebook complements the January 2015 version of DoDI 5000.02 by illustrating with selective guidance and examples how to develop and document an adequate test and evaluation (T&E) strategy. The Program Manager will use the TEMP as the primary planning and management tool for all test activities starting at Milestone A. Best practices outlined in this TEMP Guidebook should be applied to all versions of the TEMP, including the Development Request for Proposal (RFP) TEMP.

The Program Manager will prepare and update the TEMP as needed and to support acquisition milestones or decision points. The TEMP should be specific to the program and tailored to meet program needs. Accordingly, the guidance in this guidebook, in DoDI 5000.02, and in the TEMP format guide are provided to assist in developing the appropriate TEMP format and content for each program. Strict or immediate adherence to the new TEMP format is not required. Use common sense to apply the guidance to fit your program. Evaluation of TEMP adequacy is based on the TEMP's content, not the format.

Summary of the TEMP and TEMP Guidebook Format

The TEMP format has been changed as illustrated below. The previous TEMP format on the left explained in sentences and paragraphs what DOT&E required for adequacy. TEMP Guidebook 2.1 added colored callout boxes with links to the DOT&E Guidebook guidance and examples.

The new TEMP format on the right enumerates in bullets what should be considered for inclusion in each paragraph/section of the TEMP. Callouts with links to DOT&E guidance in the Guidebook 3.0 are in bold blue font.

Reliability Test Planning

<u>Operational Mission Reliability</u>: the ability of a system to perform **its required function,** under stated **environmental** and **operating conditions** and for a stated **period of time.**

- Operational testing provides the ability to assess mission reliability - testing is conducted to evaluate how systems improve mission accomplishment under realistic combat conditions.

- Ideally, adequate data on the mission reliability will be collected during operational testing, using representative users under a range of operationally realistic conditions.

- Operating characteristic (OC) curves are useful statistical tools for can be constructed to plan to length of a reliability test.

Reliability Requirements impact Test Duration

- **The duration of a test depends on the reliability requirement**
  - Pass/Fail
    - » *Probability of a fuse igniting without failure in a weapon system > 90%*
  - Time/Duration based
    - » *A howitzer must have a 75 percent probability of completing an 18-hour mission without failure.*
  - Mean time between failures (MTBF)
    - » *A howitzer mean time between failures must exceed 62.5 hours.*

Assessing Reliability Test Program Adequacy

- **Current DOT&E Guidance: OT should provide sufficient data to assess system reliability with statistical power and confidence**
      - » No default criteria is given for the level of statistical confidence and power.
- **Operating Characteristic (OC) curves can be used to determine the statistical confidence and power that a test is sized for.**
   - Required Inputs for OC curve **(what needs to be documented in the TEMP)**
      - » What is the IOT/FOT **test length/test size** described in the TEMP or test plan?
      - » What is the system's **reliability requirement**?
      - » What is the ultimate **reliability growth goal**?
   - Provide a visual of the risk trade space
      - » Consumer Risk: the probability that a bad system (i.e. below the reliability threshold) will be accepted
      - » Producer Risk: the probability that a good system (i.e. above the reliability threshold) will be rejected
- **While the statistical properties of a test do <u>not</u> determine its adequacy, they provide an objective measure of how much we are learning about reliability based on operational testing.**

Will there be enough information to adequately assess system reliability?

Operating Characteristic Curves Background

- **Required Inputs for OC curve**
  - What is the **test length/test size?**
  - What is the system's **reliability requirement**?
  - What is the desired **confidence level?**
- **Confidence level manages Consumer Risk.**
  - An 80% confidence level requires the system to demonstrate a lower 80% confidence limit on the reliability estimate from testing of at least equal to the requirement.
  - Translates into a maximum number of failures that can be witnessed during the test, called critical number of failures
  - 80% confidence equals 20% chance a system with true reliability below the requirement will be accepted

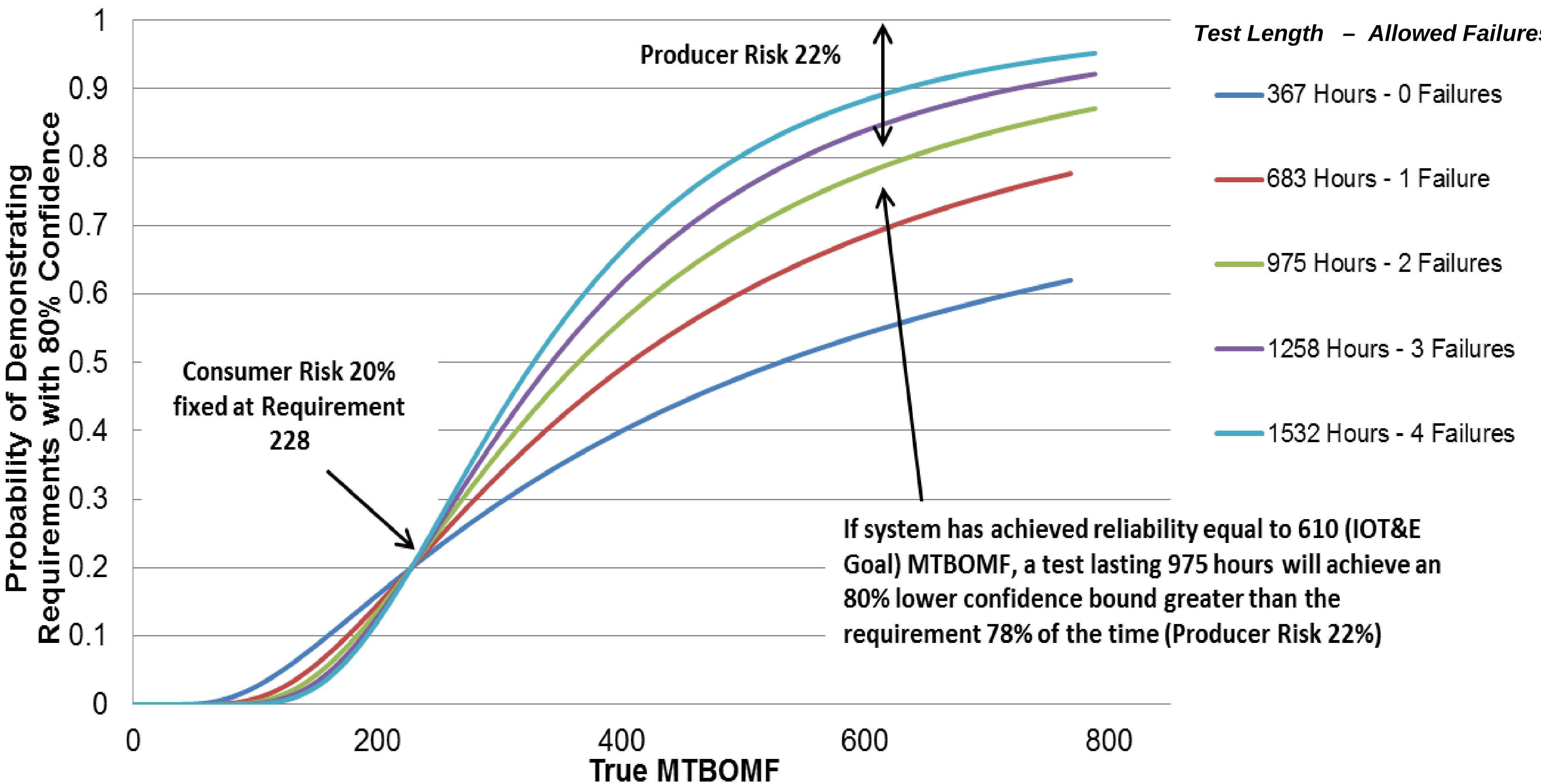
Operating Characteristic Curves Background

- **Outputs of the OC Curve**
  - Plots probability of demonstrating the reliability requirement with confidence as a function of the system under test's true reliability
    - This is the probability of exhibiting the critical number of failures, or fewer, during the test as a function of true reliability.
- **This is the power of the test**
  - Power manages Producer Risk, the higher the power the less the Producer Risk
  - Indicates the tests ability to show that a system with a true reliability higher than the requirement actually beats the requirement
- **In general, the longer the test, the higher the power for a given confidence level**

Example: Application of OC Curve

- **What reliability metrics can we apply OC Curves to?**
  - MTBOM, MTBEFF
- **Reliability Requirements:**
  - Requirement: *"The system shall have a MTBOMF that supports a 90% probability of successful completion of a 24-hour operational period"*
  - Translation: a system with a MTBOMF of 228 hours, has a 90 percent chance of experiencing zero failures in a 24 hour mission
- **Assessing the planned length of IOT&E:**
  - What risks do we take on with a planned 1,000 hours of testing in IOT&E?
- **Required Inputs for OC Curve**
  - What is the **test length/test size?**
    - » 1,000 hours of testing are planned for IOT&E
  - What is the **system's reliability requirement?**
    - » Threshold values for MTBOMF 228 hours
  - What is the desired **confidence level?**
    - » Traditionally taken to be 80% but can be varied if necessary

Constructing and Evaluating the OC Curve



The True MTBOMF is what the system needs to achieve in order to demonstrate the requirement with confidence; the reliability growth goal

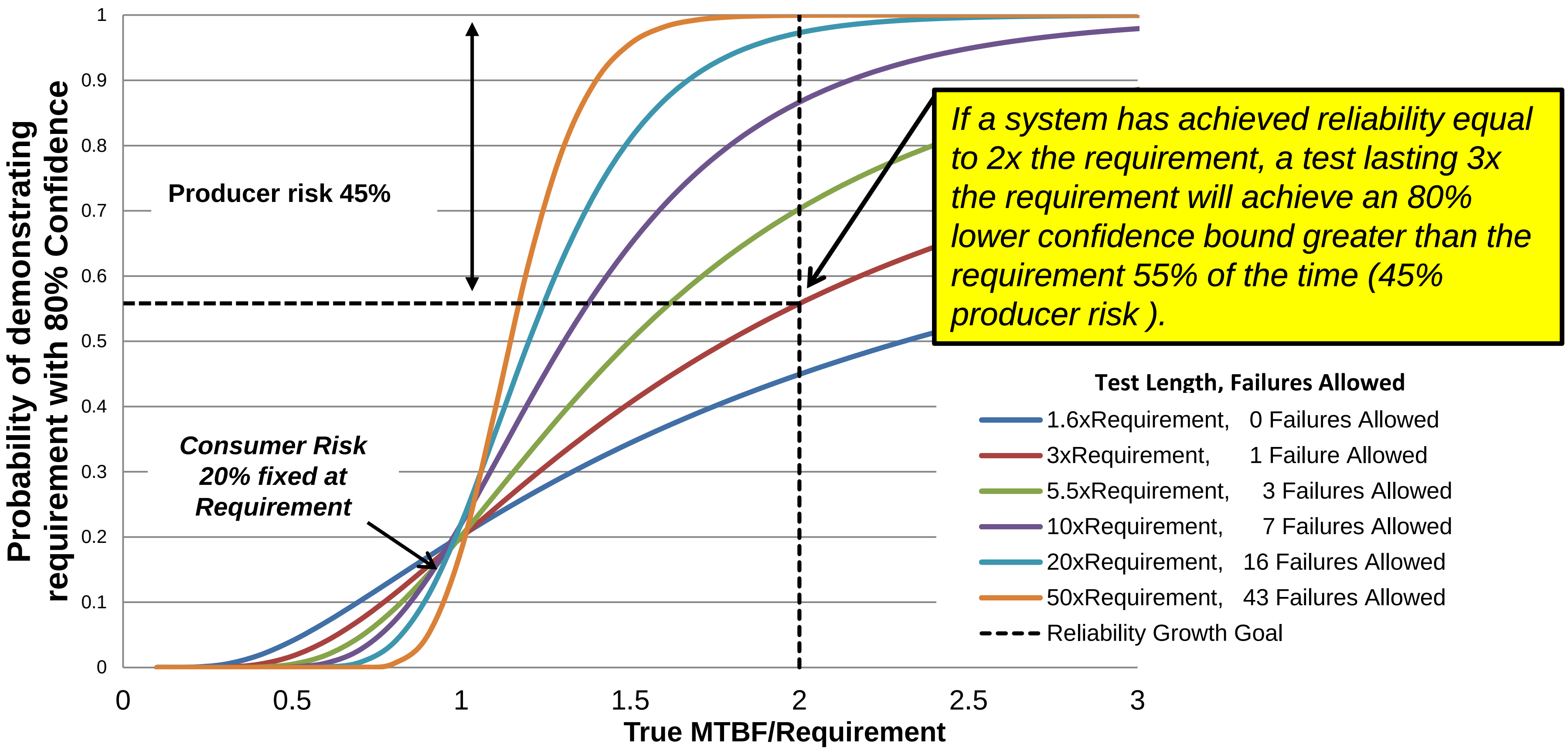
Exercise

Constructing an OC curve

Microsoft Excel Worksheet

Avoid Rules of Thumb: Test 3x the Requirement

- A Rule of Thumb should not be the strategy employed to develop or assess a reliability test plan.



Comparison across multiple OC curves helps to gauge test size as a function of allowable failures and risk.

Incorporating Additional Information

- **As it turns out, many operational tests are not statistically adequate (confidence and power) to assess requirements…**
  - Cost and Schedule Constraints
  - Requirements are not testable or not operationally meaningful
- **In most cases, there is still sufficient data to asses system reliability performance.**
  - When system reliability is substantially below the requirement, it is possible to determine with statistical confidence that the system did not meet is requirement with less testing than would otherwise be required.
  - Other sources of data can be leveraged to assess reliability…

TEMP Guidance on Incorporating Additional Information

- **If additional information (like DT data) will be used in the reliability assessment then the TEMP should specify:**
  - Conditions the data must be collected under to be acceptable for OT use
  - Methodology for scoring reliability data collected outside of an OT
  - Statistical models and methodologies for combining information (e.g. Bayesian Methods).
  - Methodology for determining an adequate operational test duration

Data from different test events should not be combined into one pool of data and used to calculate and average reliability, rather advanced analysis methodologies should be used to combine information from multiple tests.

Bayesian Reliability & Assurance Testing

Objective

- Scope an appropriately sized Operational Test (OT) using the demonstrated reliability and growth of the system under test

Demonstration Test (OC Curve Analysis)

- A classical hypothesis test, which *uses only data from single test* to assess whether reliability requirements are met - often requires an exorbitant amount of testing!
  - » OC Curve scopes the size of a Demonstration Test, balancing consumer and producer risk

Assurance Test (Bayesian Analysis)

- Leverages *information from various sources* to reduce the amount of testing required to meet a requirement.

Example: Comparison to Traditional Test Plan

Failures Allowed	Bayesian Assurance Test Miles 10% Consumer Risk 5% Producer Risk	Classical OC Curve Miles 10% Consumer Risk Producer Risk Varies
1	2,940	7,780 58% Producer Risk
2	4,280	10,645 50% Producer Risk
3	5,680	13,362 43% Producer Risk
4	7,120	15,988 37% Producer Risk
5	8,580	18,550 32% Producer Risk

A Bayesian assurance testing approach to test planning can be used to reduce test duration and control both risk criteria

Bayesian assurance test miles in table are hypothetical – only to illustrate a proof of concept

Takeaway Points

- **Reliability Growth**
  - The TEMP must provide an overview of the reliability program and the testing needed to asses and monitor reliability growth.
  - Reliability Growth Planning Curves (RGPC) should be included in the TEMP and reflect the reliability growth strategy.
  - Reliability should be measured, monitored and reported throughout the acquisition process.

- **Test Planning**
  - The duration of test depends on the reliability requirement.
  - OC Curves can be employed to visualize the risk trade space for a given test length.
  - If additional information will be used in the reliability assessment then the TEMP needs to clearly outline the source, fidelity, and methodology for combining the information.

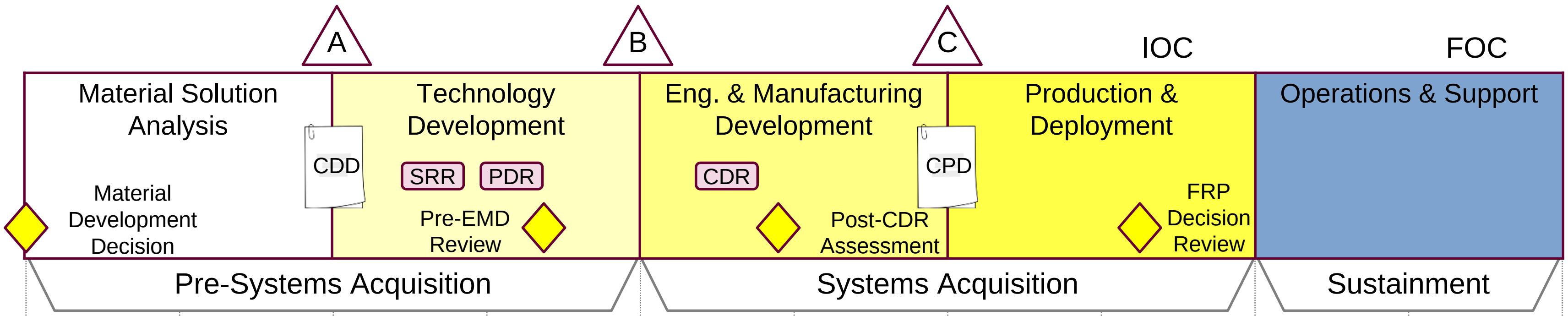
Institute for Defense Analyses
4850 Mark Center Drive • Alexandria, Virginia 22311-1882

Analysis of Reliability in Developmental Testing (DT)

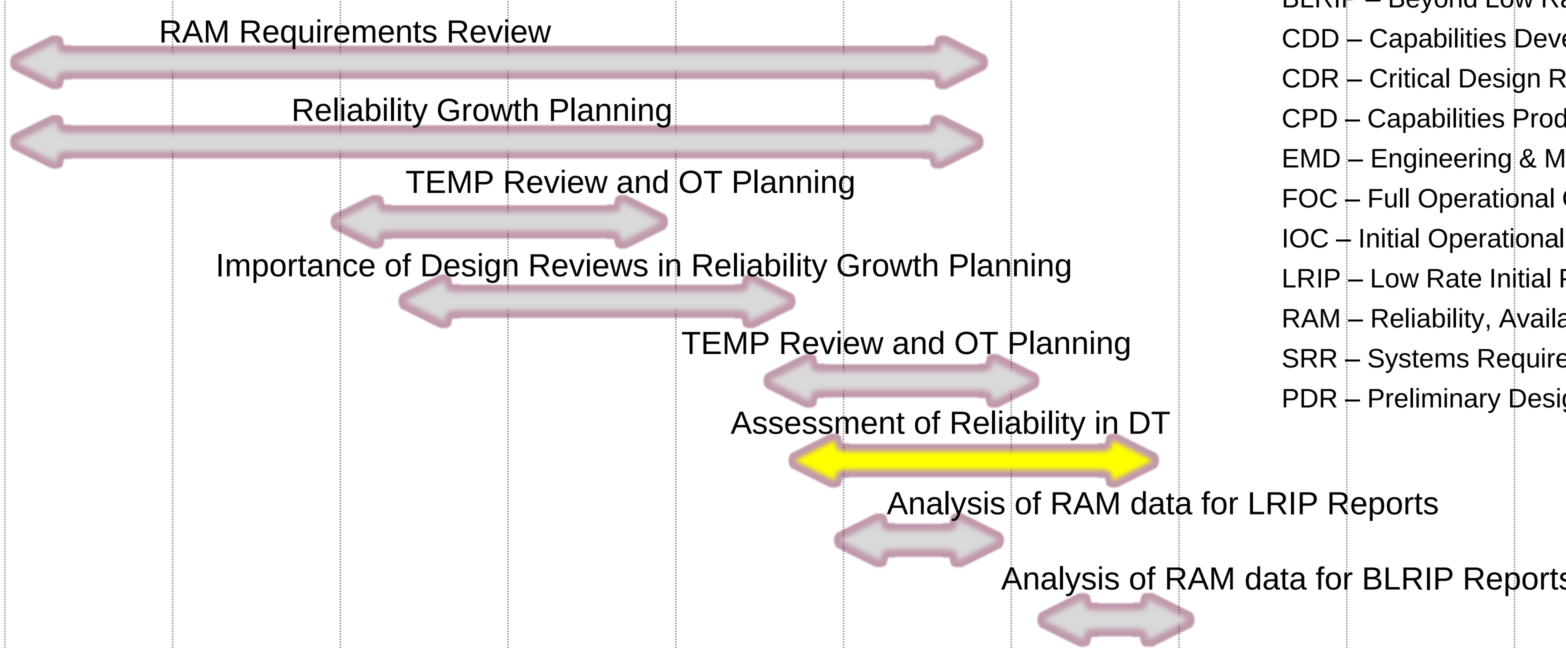
Rebecca Dickinson
2 June 2016

Timeline

System Acquisition Framework



IDA Reliability Course Topics



Acronyms:

BLRIP – Beyond Low Rate Initial Production
CDD – Capabilities Development Document
CDR – Critical Design Review
CPD – Capabilities Production Document
EMD – Engineering & Manufacturing Development
FOC – Full Operational Capability
IOC – Initial Operational Capability
LRIP – Low Rate Initial Production
RAM – Reliability, Availability, Maintainability
SRR – Systems Requirement Review
PDR – Preliminary Design Review

Overview

- **Tracking Reliability in Developmental Testing**
  - Updates to the reliability growth program
    - Reliability Growth Curves must be updated at Milestone C
  - Full system reliability tracking
  - Component level reliability tracking
  - Formal methods of reliability tracking

- **Guidance for using DT data for OT evaluations**

Importance of Developmental Test Data

- **DOT&E AO's should ensure that the Milestone C TEMP contains a reasonable reliability growth program that reflects the test results to date**
- **Key questions when assessing usefulness of developmental reliability test data:**
  - Is the test schedule/duration sufficient to surface failure modes in accordance with the reliability growth plan?
  - Is the testing being run in accordance with the Operations Mode Summary Mission Profile (OMS/MP) or Design Reference Mission (DRM)?
  - Does the program consistently document how reliability problems discovered in DT are being corrected?
  - Does the program hold assessment conferences to assess fix effectiveness of corrective actions, use reliability tracking models to assess progress, and determination if the reliability is increasing with time?
  - Do the reliability results suggest the system will be ready for OT?

5000.02: "Updated TEMPs at Milestone C will include updated Reliability Growth Curves (RGC) that reflect test results to date, any updates to the planned T&E for reliability growth, and a working link to the updated Failure mode, effects and criticality analysis (FMECA)."

Comprehensive DT Reliability Test Program: Command and Control System

- **Provides the command and control system for the Marine Corps three primary control agencies**

- **Reliability Requirement:**
  - 228 Hour Mean Time Between Critical Failure
  - Derived from 90% probability of completion for a 24 hour mission

- **Key elements of the reliability program:**
  - Program manager has made reliability a priority
  - Program has a dedicated reliability expert that attends testing, reviews all data, and tracks reliability over time
  - Planned developmental testing is 3200 hours
  - Operational realism of developmental testing is clearly documented
  - Program office conducts deficiency review boards after each test
  - Reliability tests conducted in an operational manner are also scored in an operational scoring conference according to the failure definition scoring criteria.

Proposed Testing
MS C TEMP Content

Dates	Test Event/ Location	Type of Test	#Full Systems/ Agency	Actual Full System Operational Hours	Planned Full System Operational Hours
10 Feb-7 Mar 2014	Camp Pendleton	DT-B1	1	192	200
28 May -21 Jun 2014	Camp Pendleton/ MCAS Miramar	DT-B2	3	564	600
4th Quarter FY14	Camp Pendleton	Data Fusion Testing	1	26.5	N/A
11-27 Sep 2014	MCAS Yuma, AZ	DT-B3	3	459	450
29 Sep-18 Oct 2014	MCAS Yuma, AZ	OT-B1 (OA)	3	500	450
3rd Quarter FY15	Wallops Island, VA	CEC Certification	1	-	100
3rd Quarter FY15	Camp Pendleton/ MCAS Miramar/Wallops Island	DT-C1	3		600
1st Quarter FY16	Camp Pendleton/ MCAS Miramar/MCAS Yuma, AZ	DT-C2 (IOT&E Readiness)	4	-	800
2nd Quarter FY16	MCAS Yuma, AZ	OT-C1 (IOT&E)	3	-	684

Hours are adequate to find failure modes in DT

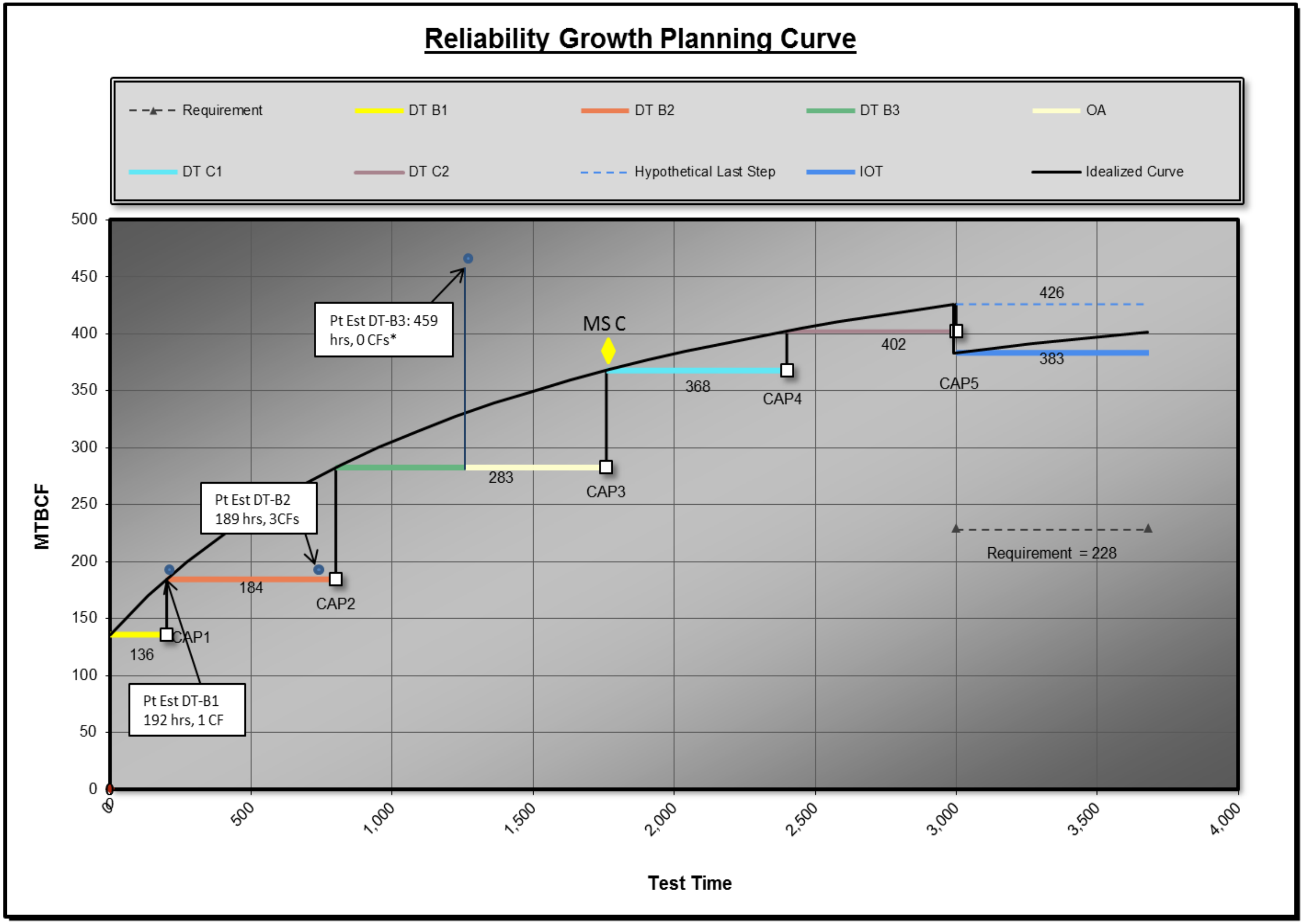
System configuration and operational scenarios are clearly documented. The majority of developmental testing is conducted using operational scenarios

Test events are space appropriately to correct discovered failures

Test Event	Hardware Production Rep (Y/N)	Non Standard Config Items (Y/N)	Software Production Representative (Y/N)	Operational Scenarios (Y/N)
DT-B1	No	Yes, Stand Alone CEP	No	No
DT–B2	No	No	No	Yes
DT–B3	No	No	No	Yes
OT-B1 (OA)	No	No	No	Yes
DT-C1	No	No	No	Yes
DT-C2 (IOT&E Readiness Event)	Yes	No	Yes	Yes
OT-C1 (IOT&E)	Yes	No	Yes	Yes

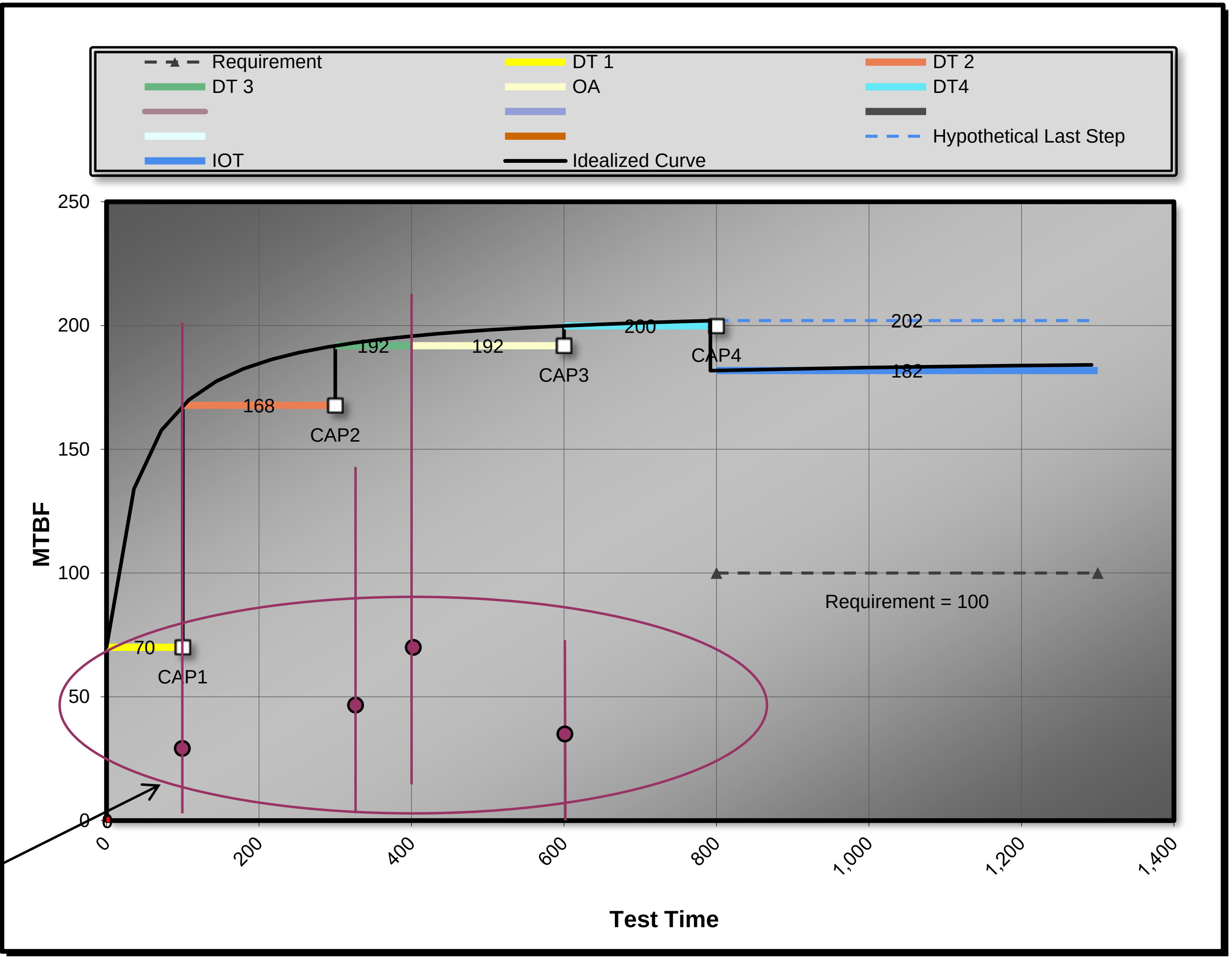
Update to Reliability Growth Curve
MS-C TEMP Content

- **Good example of one method for updating the reliability growth curve for a Milestone C TEMP**
  - Reliability for each test event is clearly documented
  - Could be improved by including confidence intervals
  - Reliability point estimates are consistent with the curve



Update to Reliability Growth Curve
MS-C TEMP Content

- **In many cases we have seen curves that do not reflect existing test data**
- **Options for MS C**
  - Update curve to reflect new initial reliability estimate – this may require:
    - » A new curve with additional corrective action periods
    - » Context on how existing failures will be fixed
  - Review requirement – what is the operational context?



Test results are not consistent with reliability growth curve!

Estimating Reliability with Limited Full System Data

- **Many programs cannot test the full system to assess reliability prior to MS-C**
  - Example: Littoral Combat Ship
- **Additionally, many systems will have limited full system level testing in an operational scenario, but lots of component level testing**
  - Example: Small Diameter Bomb II
- **In these cases it is still important to address the key developmental test considerations:**
  - Is the test schedule/duration sufficient to surface failure modes in accordance with the reliability growth plan?
  - Is the testing being conducted to address the operational context (for example temperature cycling in the lab)?
  - Does the program consistently document how reliability problems discovered in DT are being corrected?
  - Does the program hold assessment conferences to assess fix effectiveness of corrective actions, use reliability tracking models to assess progress, and determination if the reliability is increasing with time?
  - Do the reliability results suggest the system will be ready for OT?
- **However, the answers to these questions may be based on critical subsystem testing.**

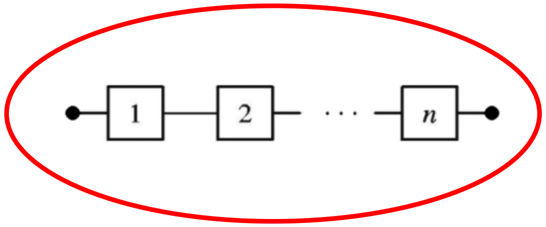
Estimating System Reliability Model: A Small Bomb

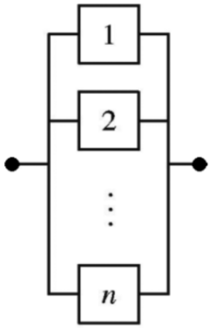
- **Reliability Key system Attribute (KSA) has a Free Flight and Material component**
  - Completion of MOT&E Free Flight Reliability (P_{FFR})= 0.80; Lot 5 P_{FFR} = 0.90
  - Material reliability MOT&E = 125 hours; Lot 5 RM = 500 hours

- **Prior to MS-C limited full-system free flight data is available to assess reliability**
  - However, each of the 24 critical components of the small bomb (ex: seeker, battery, fuze) have been tested to a greater extent, can use that information to improve the MS-C assessment of free-flight reliability.

$$R_{sys} = \prod_{i=1}^{n} R_{comp(i)}$$

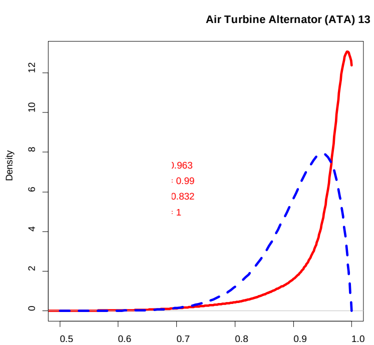
Series Reliability





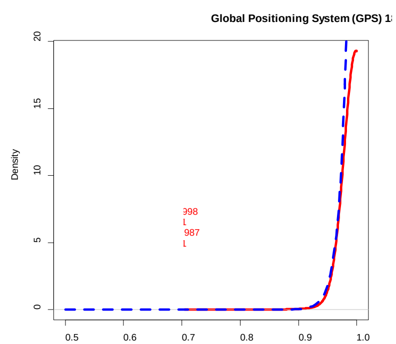
Series System

Parallel System



Air Turbine Alternator Reliability

*



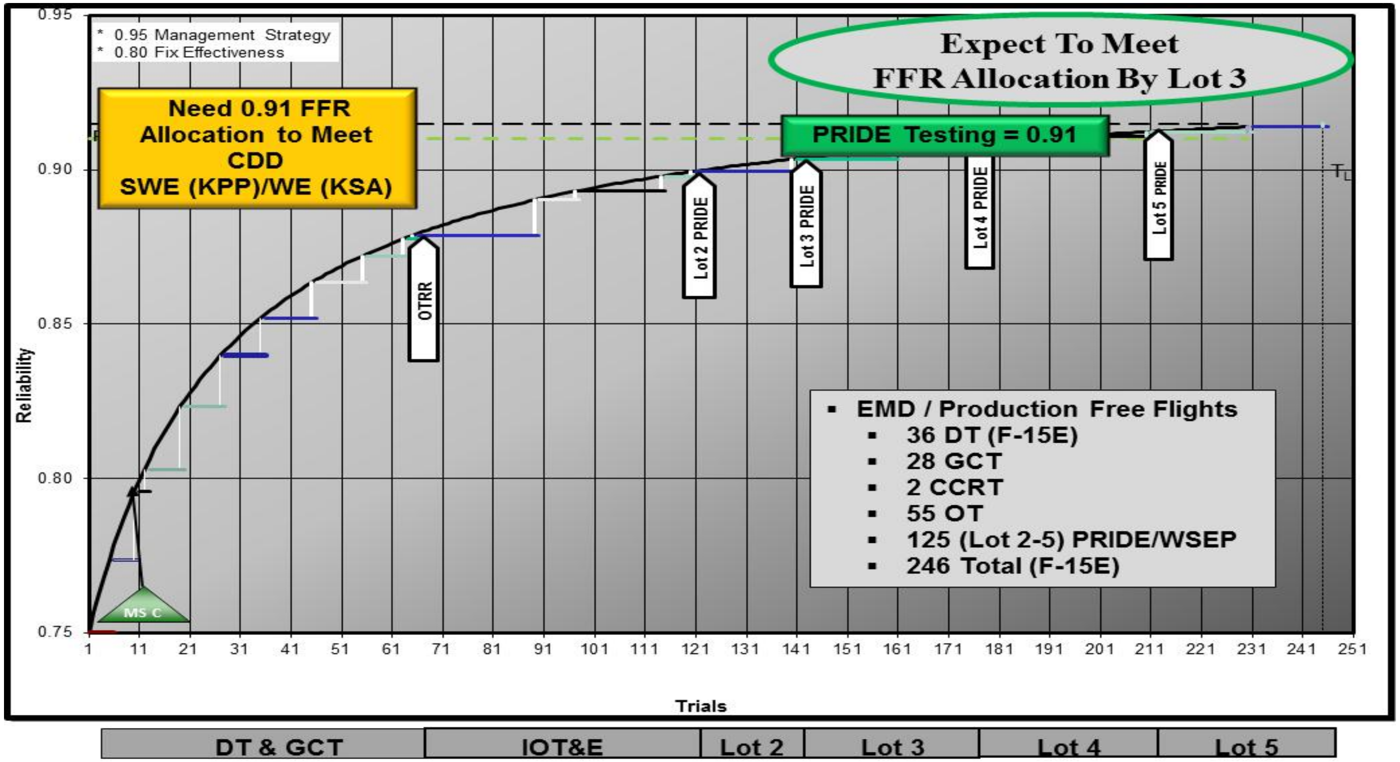
GPS Reliability

For example previous testing had shown the following success rates for two of the components:

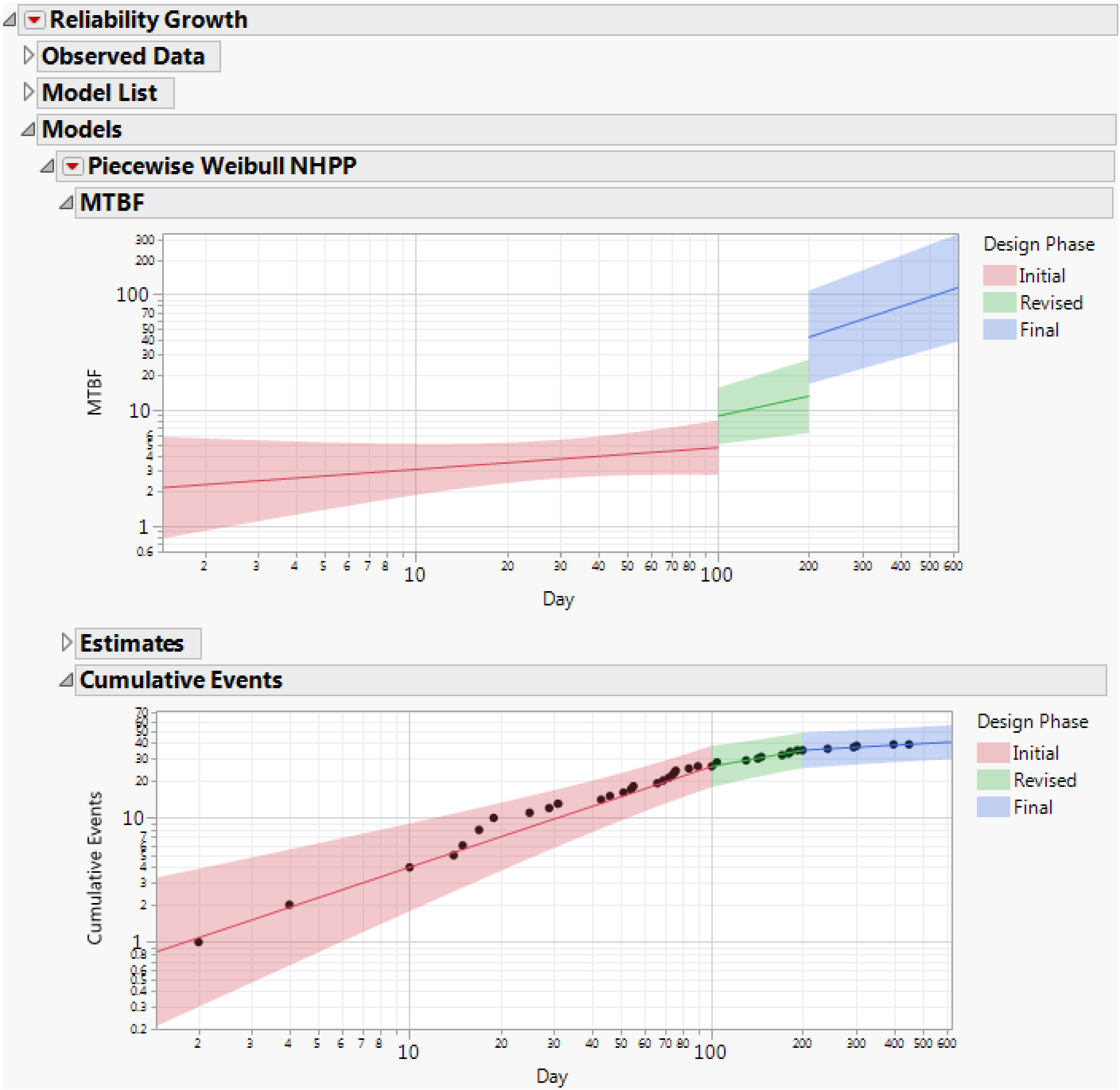
- 19/20 for Air Turbine Alternator
- 72/72 for GPS

Update to Reliability Growth Curve: MS-C TEMP Content

	Trails	Failures	Reliability	90% Upper	10% Lower
P_{FFR} Traditional Analysis	15	4	0.7333	0.88	0.54
P_{FFR} Series Model	NA	NA	0.799	0.87	0.69

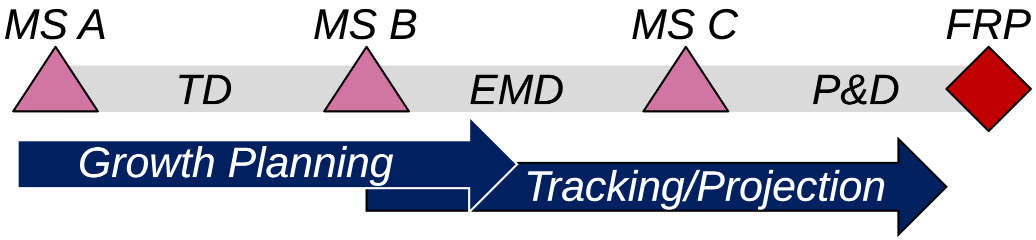


Formal Reliability Tracking



Reliability Tracking Objectives

- **Reliability Tracking Objectives**
  - Determine if system reliability is increasing with time (i.e., growth is occurring) and to what degree (i.e., growth rate)
  - Estimate the demonstrated reliability based on test data for the system configuration under test at the end of the test phase
  - Compare the demonstrated reliability to the threshold value to ascertain that reliability is growing in accordance with planned growth.



Acronyms: EMD – Engineering and Manufacturing Development
P&D – Production and Deployment
FRP – Full-Rate Production
TD – Technology Development0
MS – Milestone

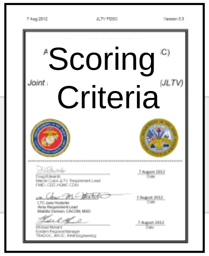
Reliability Tracking Process

Test

Collect and Review Data

- Consistency
- Omissions
- Errors
- Does it align with observations?

Score Reliability Data

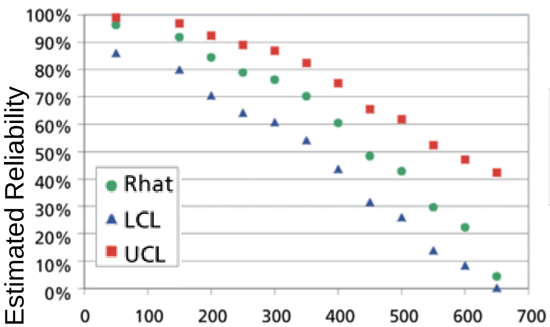


Organize Data as Appropriate

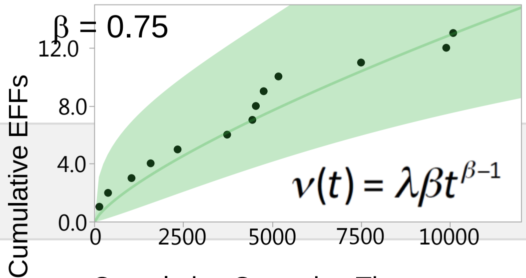
- Test phase
- Corrective action period
- Failure type (EFFs, OMFs)
- Failure mode
- System or subsystem
- Aggregate or group
- Data from legacy system(s)
- Determine what conditions might have impacted data (e.g., weather, location, etc.)

Track Reliability Growth Over Time

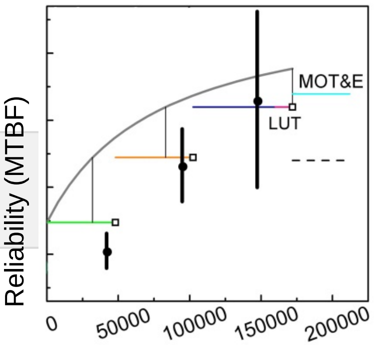
Non-Parametric Reliability Plot



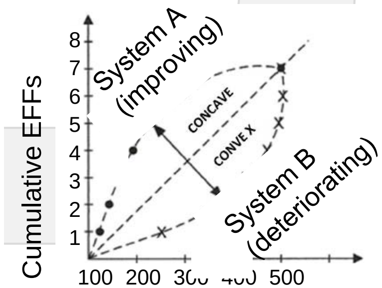
AMSAA-Crow Model Fit



Reliability demonstrated in each phase



Cumulative plot



Consider Using Tracking Data to Preform Reliability Projection Analyses

Reliability Projection Models

- AMSAA-Crow Projection Model
- Crow Extended Projection Model
- AMSAA Maturity Projection Model
- Discrete Projection Model

Reliability Growth Potential (M_{GP})

$$M_{GP} = \frac{M_I}{1-(MS)\mu_d}$$

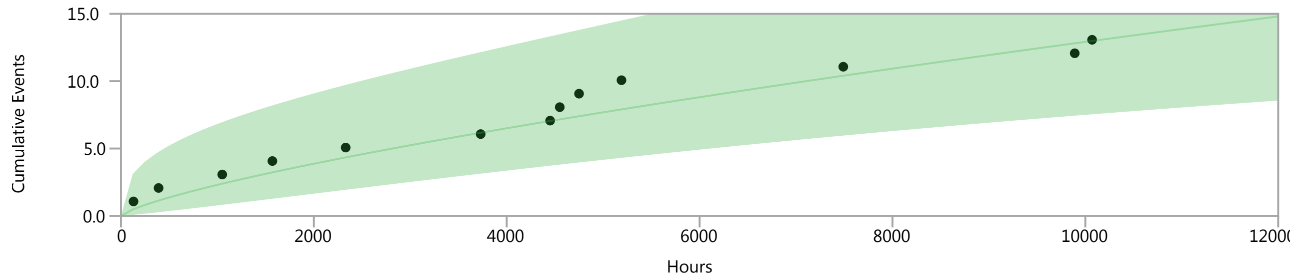
Acronyms:

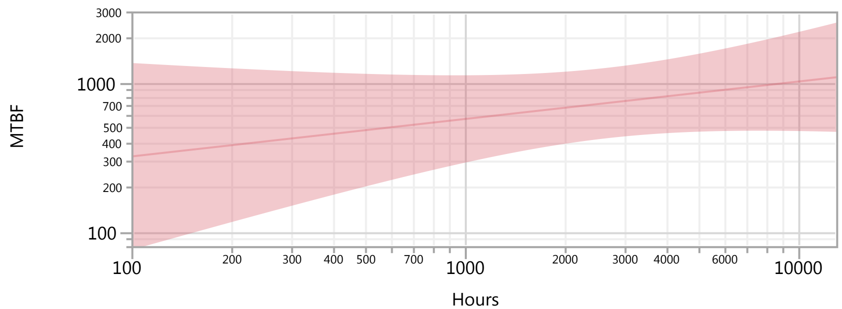
AMSAA – Army Materiel Systems Analysis Activity
OMF – Operational Mission Failure
EFF – Essential Function Failure

Reliability Tracking and Growth Potential Examples

AMSAA Crow Tracking Model (in JMP)

	Hours	Fixes
1	148	1
2	396	1
3	1051	1
4	1585	1
5	2349	1
6	3747	1
7	4456	1
8	4553	1
9	4759	1
10	5192	1
11	7492	1
12	9892	1
13	10085	1





Parameter	Estimate
lambda	0.01306317
beta	0.74878556
Reliability Growth Slope	0.25121444

Reliability Growth Potential

Reliability Measure: Mean Miles Between Operational Mission Failures			
Initial Reliability (R_i) from CAP 1	Reliability Growth Potential (R_{GP})	80% of the Reliability Growth Potential	Pre-IOT&E Reliability Goal
1,600	5,220	4, 176	3,800

Reliability growth potential
$$R_{GP} = \frac{R_i}{1-(FEF * MS)}$$

Based on *Reliability Growth Plan*

- Fix Effectiveness Factor (FEF) = 0.73
- Management Strategy (MS) = 0.95

According to page 29 of MIL-HDBK-189C, programs rarely achieve more than 80 percent of the growth potential.

Reliability Tracking Points/Recommendations

- **Once data is appropriately acquired and organized, use reliability tracking methods to:**
  – Determine if system growing reliability in accordance with reliability growth curve
  – Determine if the estimated reliability at the end of the phase is consistent with user requirements

- **If tracking and/or projection analysis indicates that the system is not growing reliability in accordance with the reliability growth curve:**
  – Update the reliability growth strategy and planning curve(s) based on more realistic inputs
  – Consider if additional resources/testing are necessary to reach goals
  – If reliability is poor, use growth potential analysis to see if it is feasible for system to reach reliability goals; if it is not feasible, system might require a redesign

Guidance on using DT data for OT evaluations

- **The conditions the data must be collected under to be acceptable for OT use.**
  – Developmental testing does not have to be conducted according to the Operational Mode Summary/Mission Profile (OMS/MP) or Design Reference Mission (DRM), but there must be a clear consideration of operational conditions in the developmental testing.
- **Use a common scoring criteria**
  – If you plan to use developmental test data for operational evaluation, developmental test reliability failures must be scored by the same methods as the operational reliability data.
- **Clearly describe the statistical models and methodologies for combining information.**
  – Data should not simply be pooled together and an average reliability calculated. The analysis should account for the conditions the reliability data were collected under to the extent possible.
- **The methodology for determining adequate operational test duration must be specified.**
  – Bayesian assurance testing can be used in place of traditional operating characteristic curves to determine adequate operational testing when prior information will be incorporated.

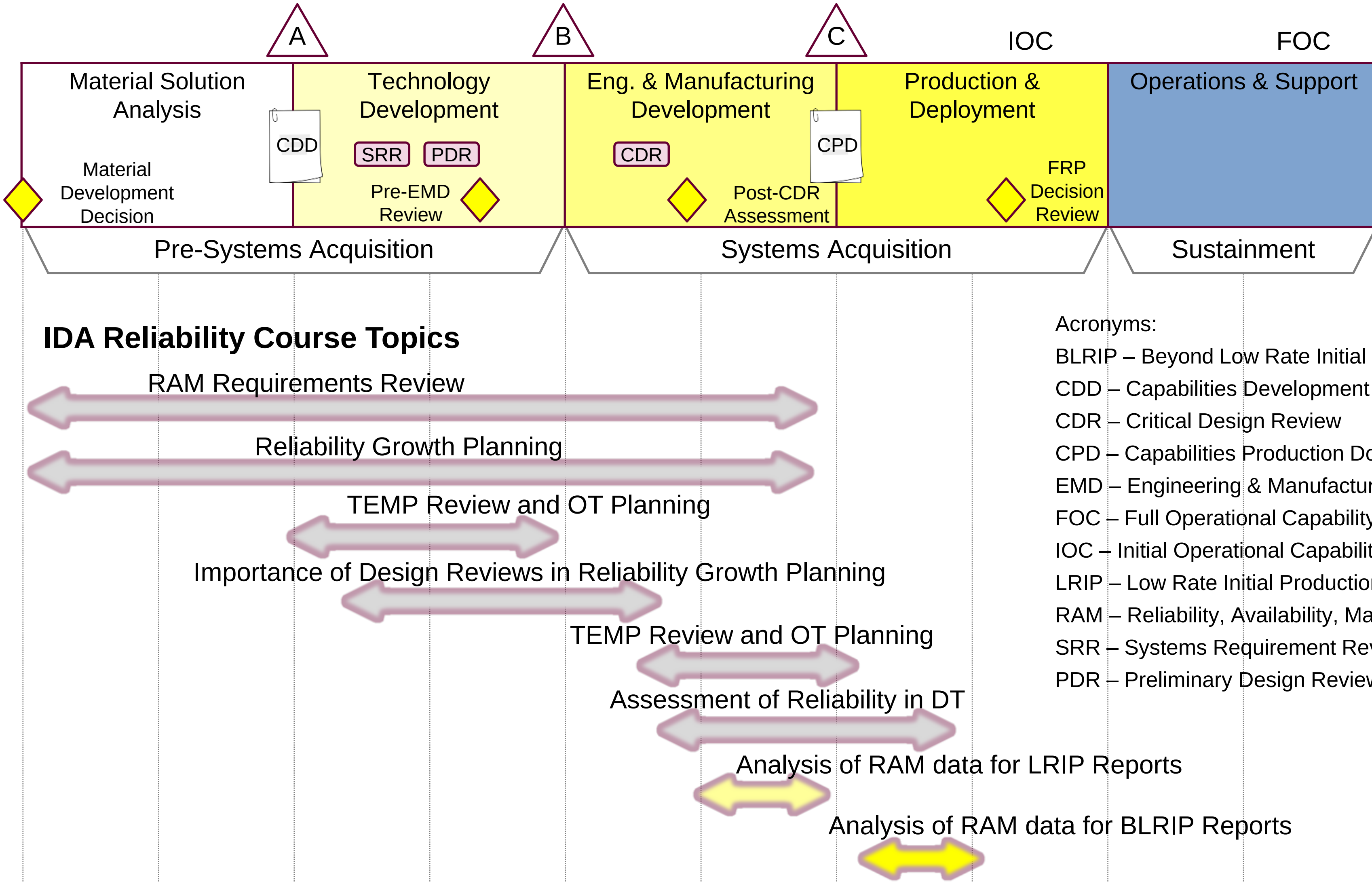
Institute for Defense Analyses
4850 Mark Center Drive • Alexandria, Virginia 22311-1882

Analysis of RAM Data for LRIP/BLRIP Reports

Matthew Avery
Rebecca Dickinson
11 December 2015

Timeline

System Acquisition Framework



Acronyms:
BLRIP – Beyond Low Rate Initial Production
CDD – Capabilities Development Document
CDR – Critical Design Review
CPD – Capabilities Production Document
EMD – Engineering & Manufacturing Development
FOC – Full Operational Capability
IOC – Initial Operational Capability
LRIP – Low Rate Initial Production
RAM – Reliability, Availability, Maintainability
SRR – Systems Requirement Review
PDR – Preliminary Design Review

Outline

- **Reporting on Reliability**
  - Point & interval estimation
  - Comparisons with legacy systems
  - Comparisons against requirements
- **Reliability Models**
  - Exponential Distribution
  - Other models (Weibull, LogNormal, …)
  - Nonparametric methods (Bootstrap)
- **Scoring Reliability**
- **Leveraging Information from Multiple Test Periods**
  - Can we combine data across OT events?
  - Can we capitalize on DT data?
- **Qualitative Assessment**
  - Identifying drivers of reliability
- **Summary**

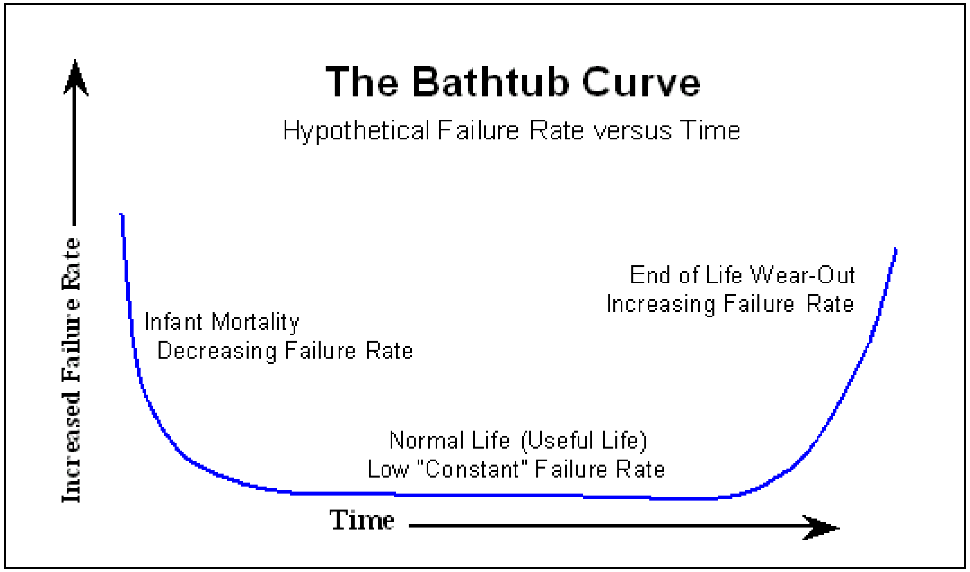
Reporting on Reliability

- **Top level assessment**
  - Was the system reliable?
  - In the first sentence/paragraph in the Operational Suitability section
- **What was the system's demonstrated reliability?**
  - Point estimate
  - Confidence interval
- **Did the system meet its requirements?**
  - Is there a statistically significant difference?
  - Is the difference meaningful in operational context?
- **How does the system's reliability compare to legacy system?**
  - Did an upgrade improve reliability? Degrade reliability?

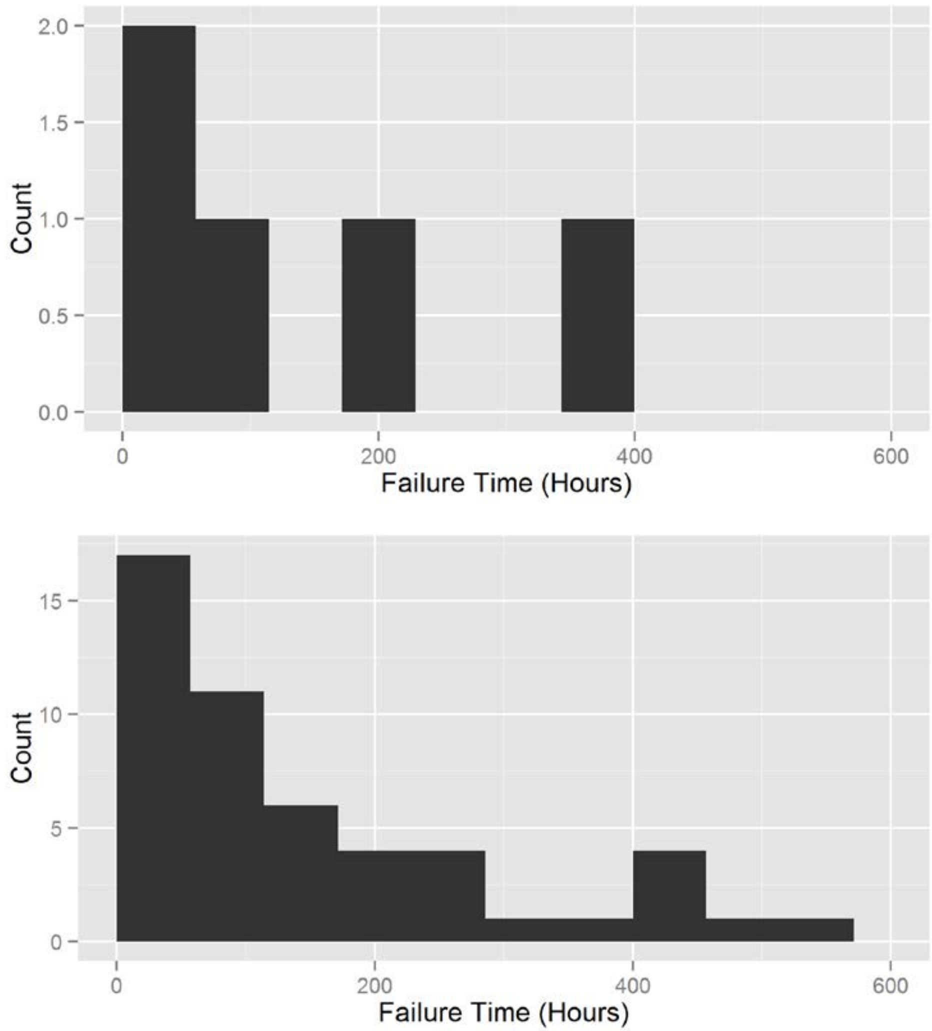
Is the system sufficiently reliable to successfully conduct its mission?

Estimating System Reliability

- **Failure times are the primary metric used to evaluate reliability**
- **Average of all times between failure = Mean Time Between Failures (MTBF)**
  - Easy to calculate
  - Requirements often given in terms of MTBF
  - Implies assumption of constant failure rates
- **Failure rates are not always constant!**
  - Median failure time provides more direct measure of frequency of failures
- **Different assumptions require different analyses**



Interval Estimation



- **Requirement: 100 MFHBSA**
- **Operational Assessment (OA):**
  - 723 hours
  - 5 failures observed
  - 144.6 MFHBSA
- **Initial Operational Test (IOT):**
  - 7052 hours
  - 49 failures observed
  - 143.9 MFHBSA

Both versions demonstrated the requirement, but we have more information about one than the other.

MFHBSA – Mean Flight Hours Between System Aborts

Quantifying Uncertainty

- **We can quantify our certainty about point estimates (such as MFHBSA) using confidence intervals**
  - Provides range of plausible values
  - Shows how sure we are about system reliability
  - Helps us evaluate risk that system meets requirement

- **Increment 1:**
  - 723 hours
  - 5 failures observed
  - 144.6 MFHBSA
  - 80% CI: **(77.9,297.2)**

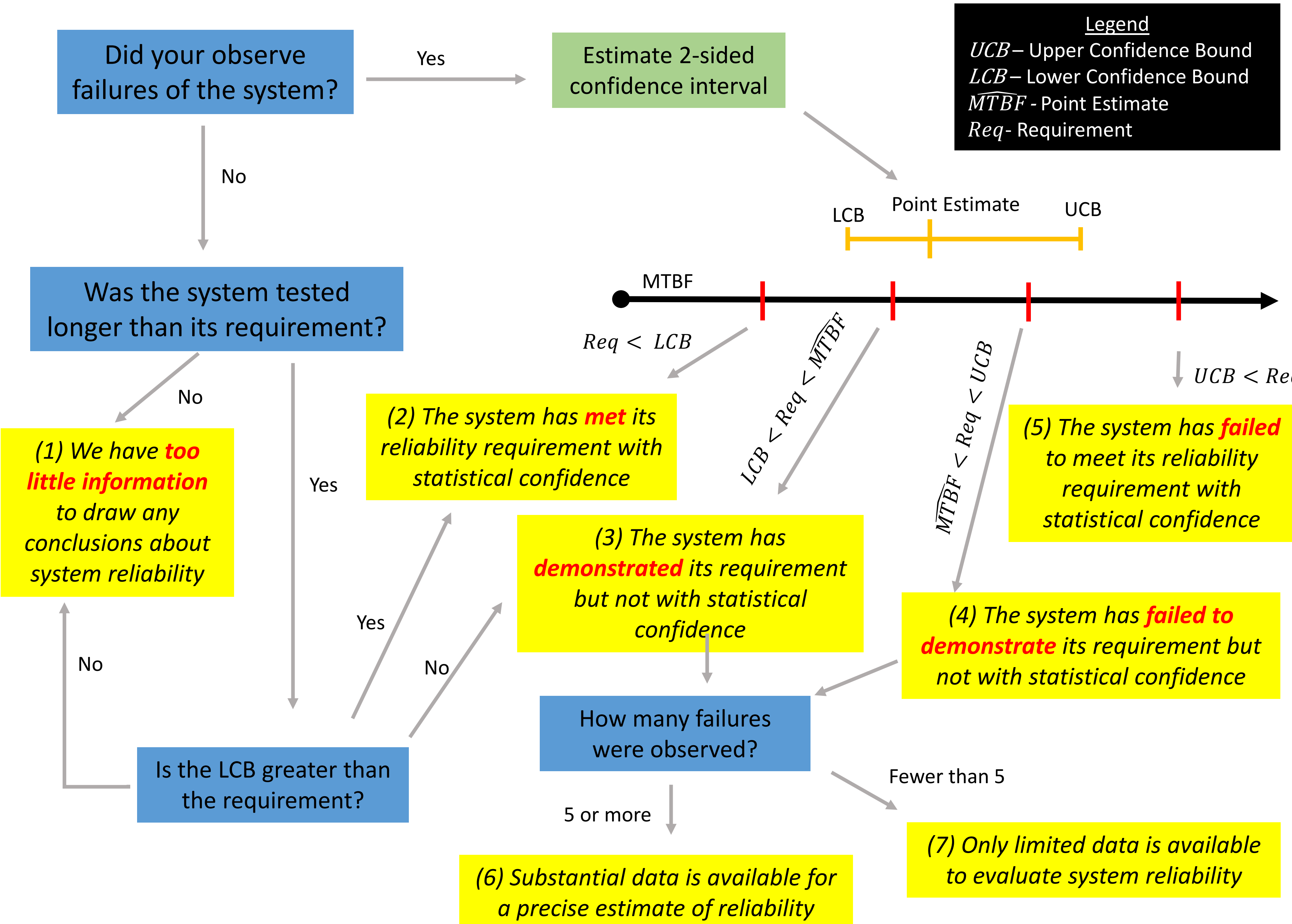
- **Increment 2:**
  - 7052 hours
  - 49 failures observed
  - 143.9 MFHBSA
  - 80% CI: **(119.0,175.1)**

Confidence Intervals for Exponential Failure Times

$$\frac{2T}{\chi^2\left(1-\frac{\alpha}{2}, 2(r+1)\right)} < MFHBSA < \frac{2T}{\chi^2\left(\frac{\alpha}{2}, 2r\right)}$$

T: Total Test Time
χ^2: Critical Value of a Chi-Squared distribution
r: Observed number of failures
α: 1-confidence level (for 80%, $\alpha = 0.2$)

MFHBSA – Mean Flight Hours Between System Aborts


Did your observe failures of the system?
Yes
Estimate 2-sided confidence interval
Legend
UCB – Upper Confidence Bound
LCB – Lower Confidence Bound
$\widehat{MTBF}$ - Point Estimate
Req - Requirement
No
LCB
Point Estimate
UCB
MTBF
Was the system tested longer than its requirement?
$Req < LCB$
$LCB < Req < \widehat{MTBF}$
$\widehat{MTBF} < Req < UCB$
$UCB < Req$
No
(1) We have too little information to draw any conclusions about system reliability
(2) The system has met its reliability requirement with statistical confidence
(5) The system has failed to meet its reliability requirement with statistical confidence
Yes
(3) The system has demonstrated its requirement but not with statistical confidence
Yes
No
No
(4) The system has failed to demonstrate its requirement but not with statistical confidence
How many failures were observed?
Is the LCB greater than the requirement?
Fewer than 5
5 or more
(6) Substantial data is available for a precise estimate of reliability
(7) Only limited data is available to evaluate system reliability

IDA

Comparisons to Legacy Systems and Thresholds

- **Does the system improve on the reliability of the legacy system?**
  - Test legacy system in side-by-side comparison
  - Use past deployment data from legacy system
    - » How closely does OT environment mimic deployment? OMS/MP?
  - Legacy system test data
    - » How closely does new test environment mimic legacy testing?

- **Did the system meet its threshold?**
  - Point estimate?
  - Lower bound of confidence interval?

*"The demonstrated system reliability was [**better**/**worse**] than the requirement*, and the difference [**was**/**was not**] statistically significant."*

**When evaluating reliability prior to the IOT, demonstrated reliability should also be compared to the reliability growth curve to determine of programs are on track to eventually meet their requirement.*

Reporting Reliability in the Mission Context

- **Reliability Requirement: "The Amphibious Vehicle shall have a 0.77 probability of completing any single one of the scenarios described in the OMS/MP"**
  - Scenarios described last at most 18 hours → 69 hours MTBSA
  - Hypothetical result from testing: 55.4 (48.6, 63.4) hours MTBSA
  - "The probability of the Amphibious Vehicle completing an 18 hour mission without experiencing a system abort is 0.72 (0.69, 0.75)."

"Over the course of the 4-day mission described in the OMS/MP, a Reinforced Rifle Company supported by 21 vehicles would expect to experience 27.3 system aborts vice 21.9 system aborts if the Amphibious Vehicle had achieved its requirement."

Provide interpretation of demonstrated system reliability in the context of the mission

MTBSA – Mean Time Between System Aborts
OMS/MP – Operational Mode Summary/Mission Profile

Reliability Models: Modeling Failure Times

Statistical models allow us to:

- *Estimate overall failure rates*
- *Quantify uncertainty through confidence intervals*
- *Compute probability of completing a mission without a failure*
- *Compare system reliability against a threshold or a legacy system*

- **Approaches discussed previously rely on statistical models**
  - When reporting the MTBF (= Total Time / Total # of Failures) we are inherently assuming that failure time data follow an exponential distribution!

- **To ensure estimates of reliability are accurate, choosing the correct model is crucial**
  - Exponential
  - Weibull
  - Nonparametric approaches

The Exponential Distribution

- **Constant Failure Rates**
  - No "infant mortality"
  - No "wear out"
  - Should always attempt to validate these assumptions with test data

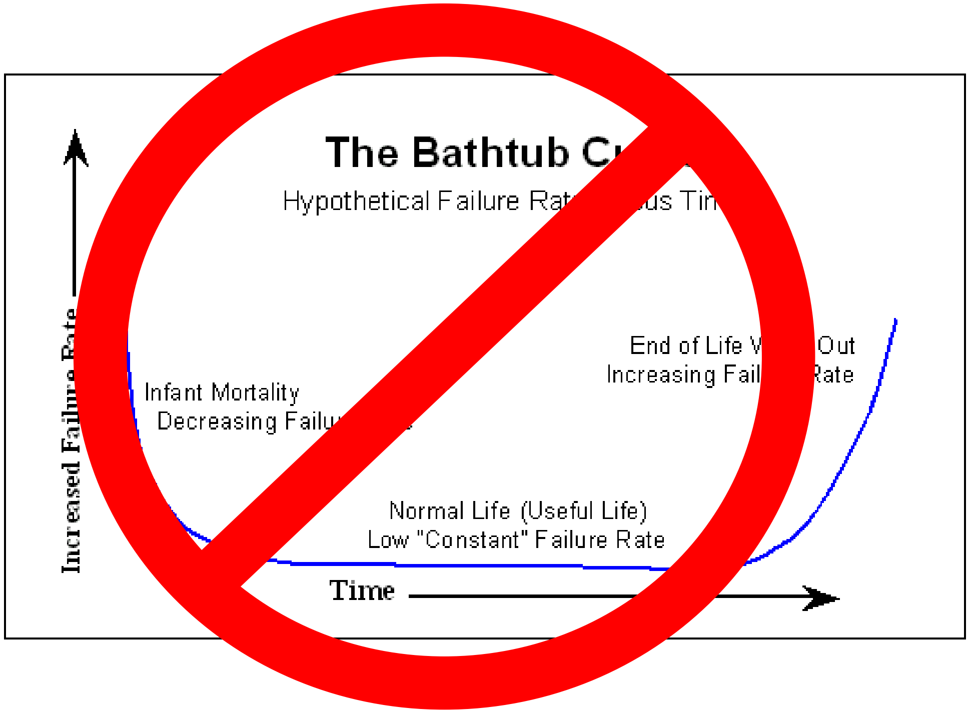
Exponential Distribution

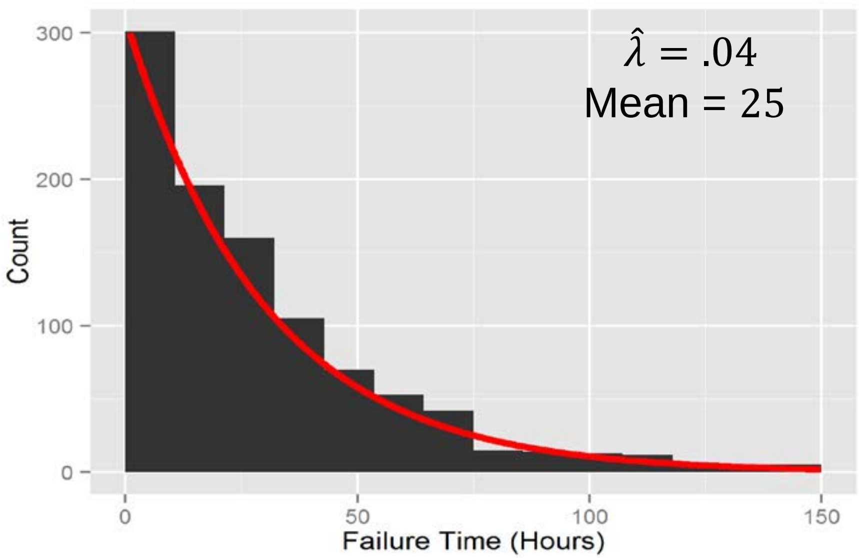
$$f(t) = \lambda e^{-\lambda t}$$

λ : the rate parameter

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}$$

Mean: $\frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} t_i}{n} = \text{MTBF}$





Why Use the Exponential Distribution?

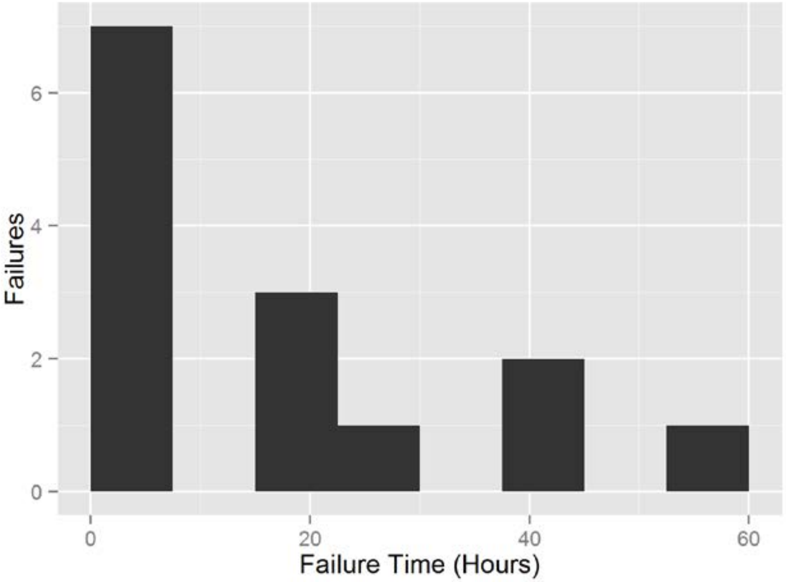
- **Intuitive, Traditional, Convenient**
  - Constant failure rates make interpretation easier
  - 1982 DoD Reliability Primer showed the calculations for mean and confidence interval
  - Someone put it in an excel spreadsheet
- **"Mean Time Between Failure"**
  - This measure makes the most sense in the context of exponential distribution
  - For alternative models (lognormal, Weibull), measures like median failure time make more sense
- **Minimal data collection requirements**
  - Total number of hours/miles
  - Total number of failures

$$MTBF = \frac{Total\ Hours/Miles}{Total\ Failures}$$

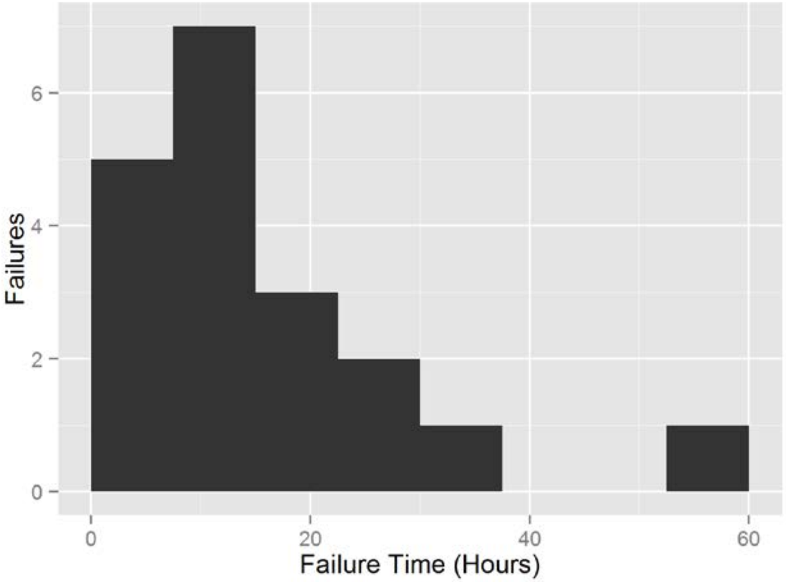
Downside of Exponential

- **Assumptions may be invalid**
  - Wider confidence intervals
  - Mis-represent system reliability
    - Over-estimate frequency of early failures

Exponential Distribution?

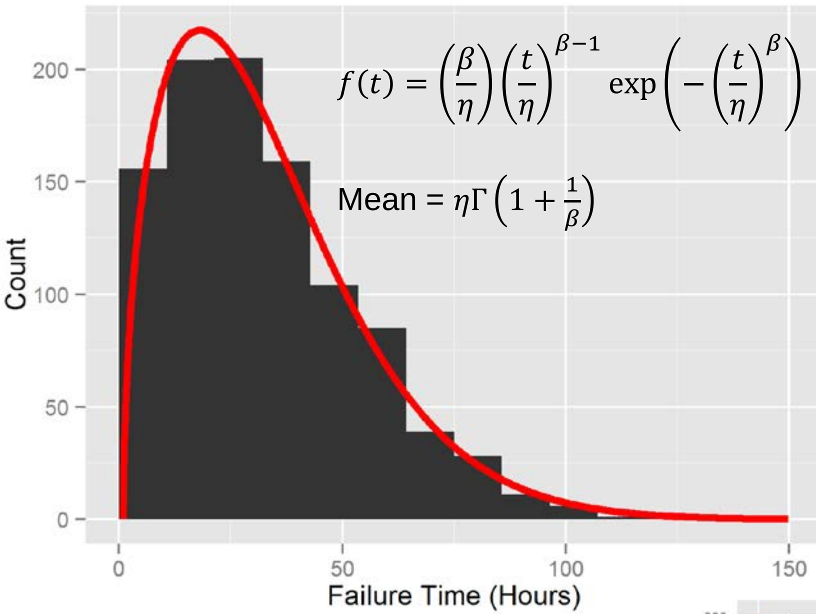


Exponential Distribution?



Alternatives to Exponential Distribution

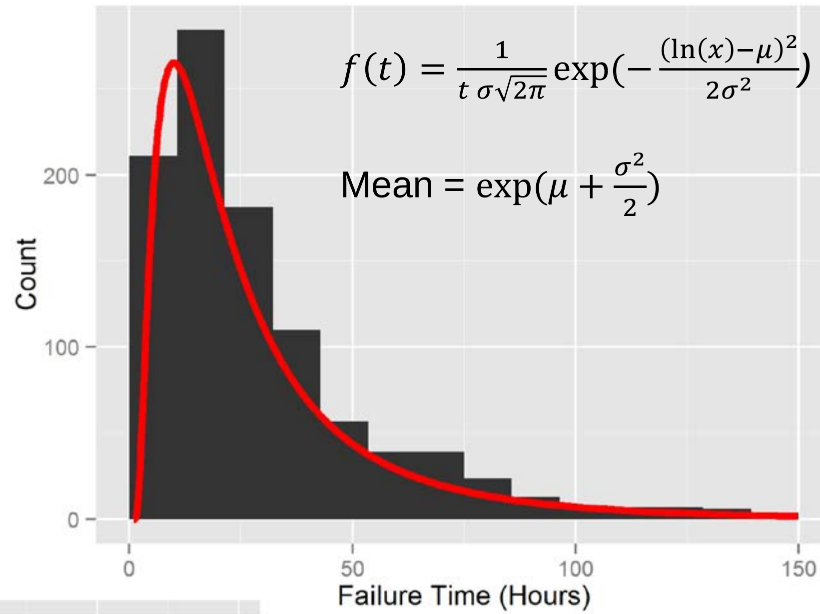
Weibull Distribution



$$f(t) = \left(\frac{\beta}{\eta}\right)\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\eta}\right)^{\beta}\right)$$

Mean = $\eta\Gamma\left(1 + \frac{1}{\beta}\right)$

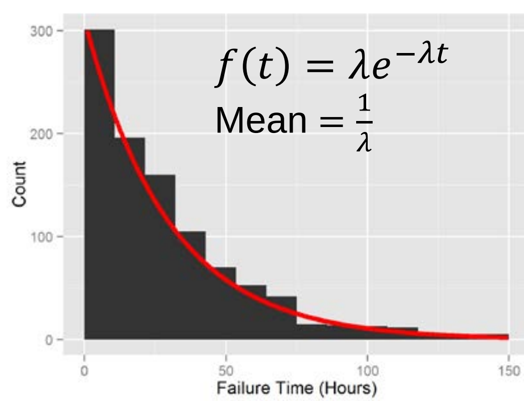
Lognormal Distribution



$$f(t) = \frac{1}{t\,\sigma\sqrt{2\pi}} \exp(-\frac{(\ln(x)-\mu)^2}{2\sigma^2})$$

Mean = $\exp(\mu + \frac{\sigma^2}{2})$

Exponential Distribution

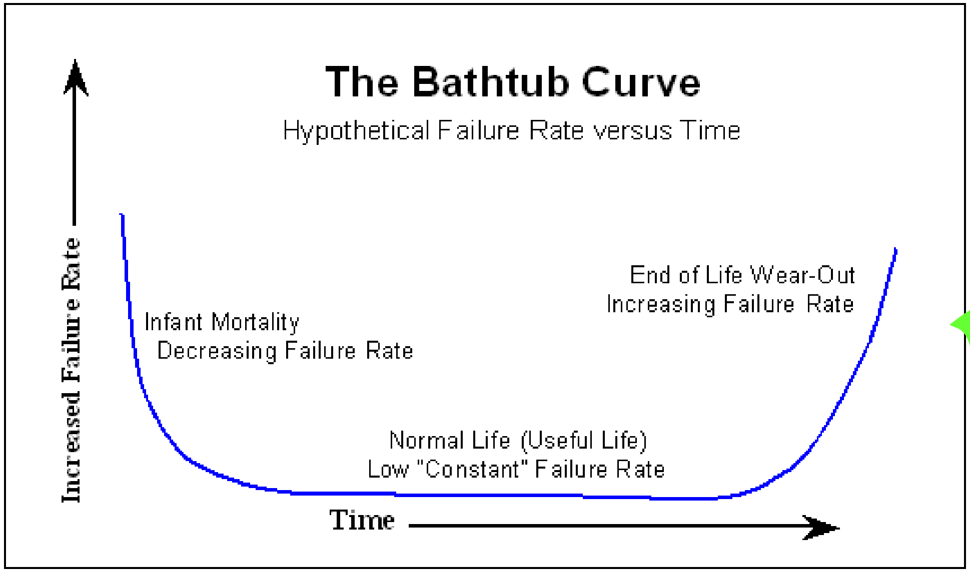


$$f(t) = \lambda e^{-\lambda t}$$

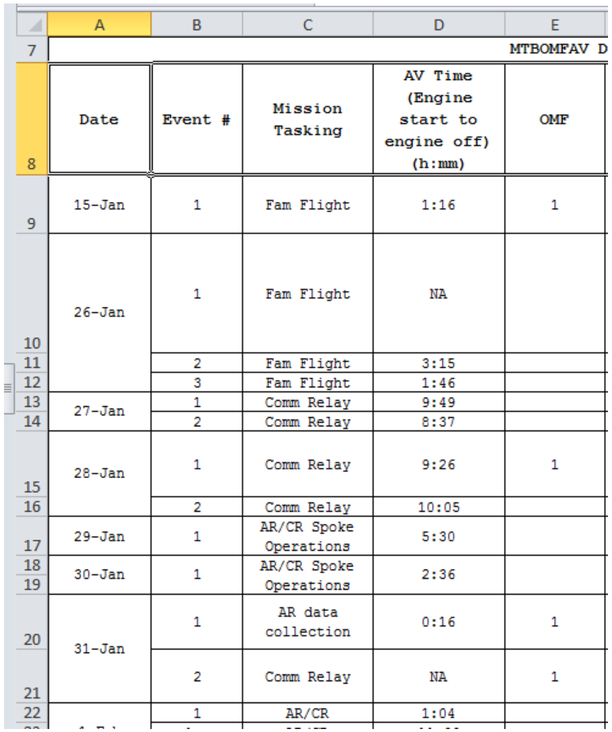
Mean = $\frac{1}{\lambda}$

Weibull and Lognormal

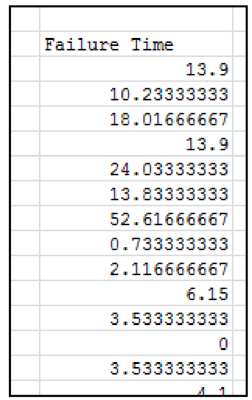
- **Multiple parameters allow for both infant mortality and wear-out at end of life**
  - Better fit of the data
- **Need time between each failure**
  - Requires planning prior to test to ensure adequate data collection



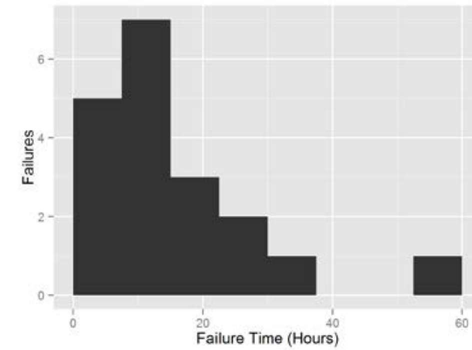
Fitting Failure Time Data

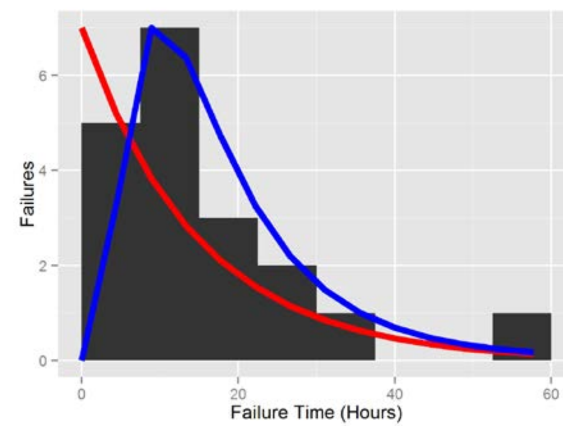


	A	B	C	D	E
7					MTBOMFAV D
8	Date	Event #	Mission Tasking	AV Time (Engine start to engine off) (h:mm)	OMF
9	15-Jan	1	Fam Flight	1:16	1
10	26-Jan	1	Fam Flight	NA	
11		2	Fam Flight	3:15	
12		3	Fam Flight	1:46	
13	27-Jan	1	Comm Relay	9:49	
14		2	Comm Relay	8:37	
15	28-Jan	1	Comm Relay	9:26	1
16		2	Comm Relay	10:05	
17	29-Jan	1	AR/CR Spoke Operations	5:30	
18 19	30-Jan	1	AR/CR Spoke Operations	2:36	
20	31-Jan	1	AR data collection	0:16	1
21		2	Comm Relay	NA	1
22		1	AR/CR	1:04	

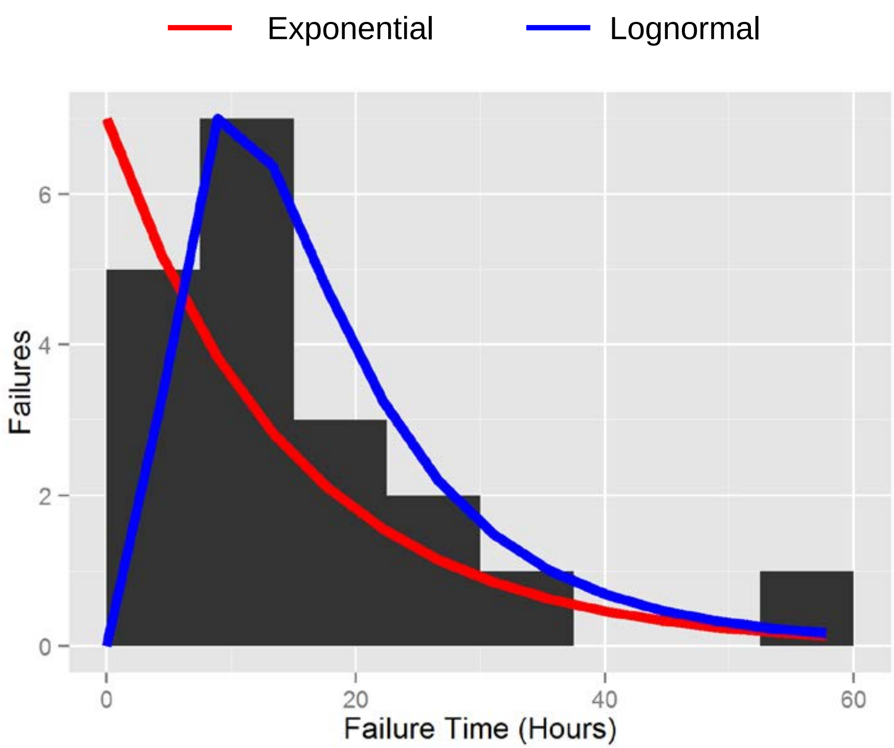


Failure Time
13.9
10.23333333
18.01666667
13.9
24.03333333
13.83333333
52.61666667
0.733333333
2.116666667
6.15
3.533333333
0
3.533333333





Checking Model Assumptions

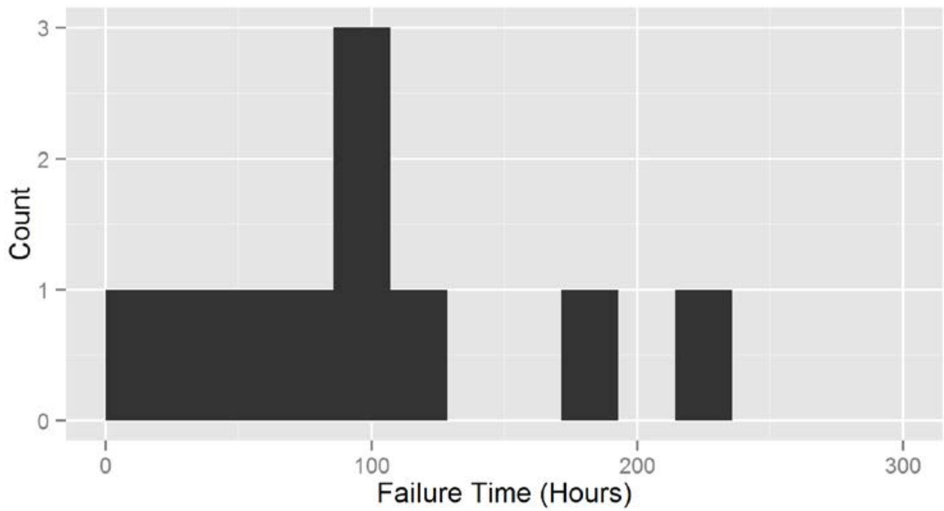


Model	Likelihood	AIC	BIC
Exponential	16.24	6.50	10.89
Lognormal	18.16	3.54	6.02

- **Compare plotted data to estimated model**
- **Goodness of fit criteria**
  - Likelihood
  - AIC/BIC

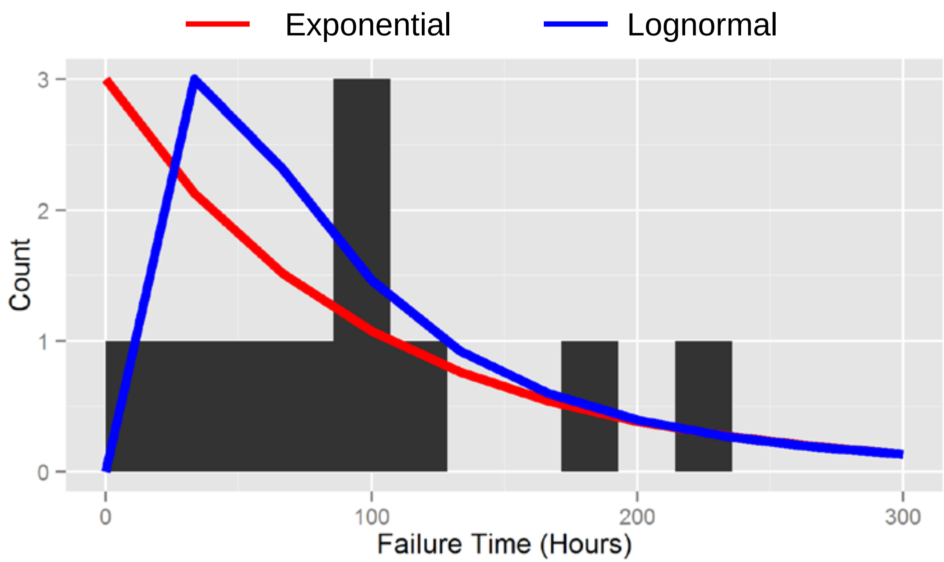
Moving Away From Assumptions

- **Observed 10 failures over 970 hours of testing:**



Moving Away From Assumptions

- **Observed 10 failures over 970 hours of testing:**



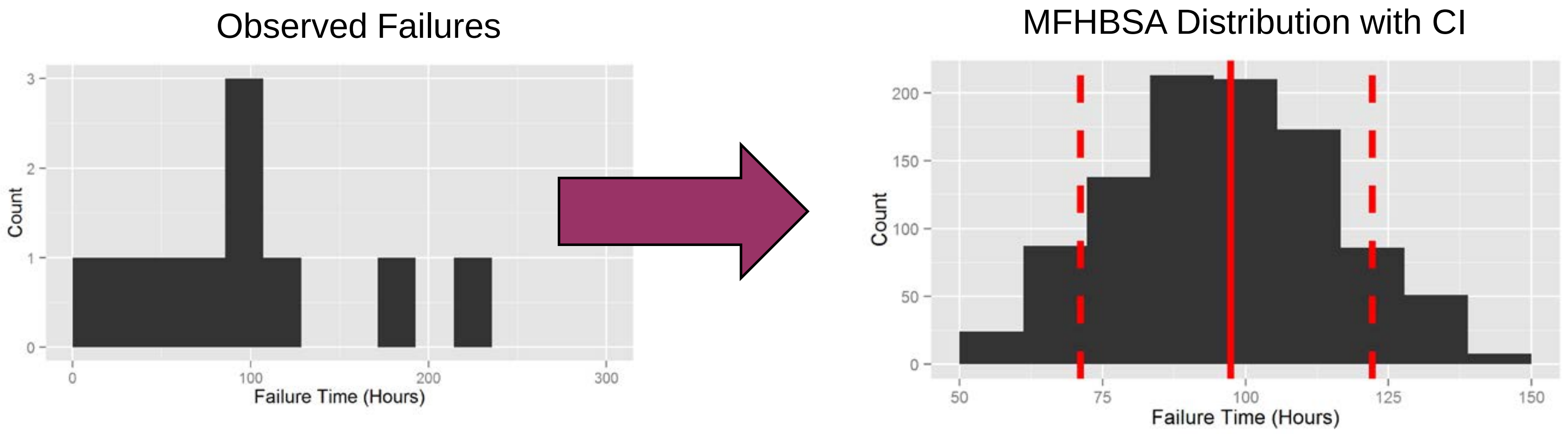
- **These models don’t appear to fit the data well**

- **Alternative methods that don’t assume a particular distribution can be used to generate uncertainty estimates**

Bootstrapping

- **Regardless of what the data is, allows you to provide uncertainty estimates**
- **Resampling based exclusively on observed data shows what "could have been"**

- *Need failure times vice aggregation*
- *Can't bootstrap with too few (<7) data points*
- *Less precise than parametric approach*



Reliability Scoring

DOT&E's evaluation of reliability is not constrained by the scoring decisions of Operational Test Agencies (OTA) or Program Managers (PM)

- **DOT&E's reliability scoring should be independent**
  - DOT&E participates in reliability scoring conferences for many programs, but DOT&E's evaluations are never bound by their decisions
- **DOT&E is not a signatory to the FDSC**
  - Failure Definition Scoring Criteria (FDSC) are developed by the services for their evaluations of systems
  - Definitions provided in FDSCs can vary substantially from service to service and may even be different for similar programs within the same service
  - DOT&E's definition of Operational Mission Failures (OMF) or system Aborts (SA) may be different from the FDSC
- **Disagreements between DOT&E scoring and OTA scoring should be highlighted in test reports, since these differences will lead to different results in reliability evaluation and estimates of failure rate**

Determining Whether Test Time Should Count for Reliability

DOT&E will make independent decisions regarding what constitutes score-able test time

- **System operating time should only accrue if the system is being operated in realistic conditions**
  - OMS/MP, CONOPS, or design reference missions may be used as resources to determine the stress expected on a system over time
- **Passive/overwatch time**
  - OMS/MP may specify that electronic systems will operate for a certain percentage of the time
    - » Anti-Tank Vehicle (ATV) turret is only credited with 37.5 hours of operating time over a 48-hour mission in the OMS/MP
- **Environmental stresses**
  - OMS/MP may specify, for example, the type of terrain the system is expected to endure
    - » ATV OMS/MP specifies:

Terrain Type	Miles (%)
Primary Road	10
Secondary Road	20
Trail	30
Cross Country	40

CONOPS – Concept of Operations
OMS/MP – Operational Mode Summary/Mission Profile

Using All of the Data Available

- **Combining data across OT events has many advantages**
  - More information to use in system assessment
  - Alleviates pressure to size test based on reliability requirement
  - Greater efficiency in testing
  - May not be possible to adequately assess some requirements through Initial Operational Test (IOT) alone
  - In some cases, may even incorporate DT data

- **However, the approach used to combine data must account for differences in data collection across phase**
  - How much does the system change from one phase to the next?
  - Is the system being operated by warfighters with the same level of training in all phases?
  - Are the operating conditions the same across test phase?
  - Was data collection/scoring conducted in the same way across different phases?

Combining Operational Test Data: The Gray-Lewis Test

- **Compare failure rates across two periods of testing**
  - Different periods of time
  - Different system configurations (be careful with this one)
  - Different test venue

- **The Grey-Lewis test is a formal statistical hypothesis test for comparing failure rates (λ)**

$$H_0: \lambda_{IOT} = \lambda_{FOT}$$
$$H_1: \lambda_{IOT} \neq \lambda_{FOT}$$

  - If failure rates are very different, the test periods should be evaluated separately
  - If failures rates are roughly similar, we can combine the data for analysis

- **CAUTION:**
  - Best used when dealing with operational test data only
  - No way to get partial credit
  - Will only detect large deviations when the individual test durations are small
  - The test cannot prove that you *can* combine information
  - The test can only prove that you *cannot* combine information

Example: UAS IOT&E Reliability

- **UAS is a small, tactical UAV**
  - Five air vehicles, ground control station with four operator work stations, launcher, recovery system, other surface components
  - IOT&E conducted January through December, 2014
    - » 29 Palms
    - » Camp Lejeune
    - » Aboard USS Anchorage

- **Test Event Similarities**
  - Same test system
  - Same test personnel

- **Differences**
  - Surface components/configuration different aboard ship and on ground
  - Environment (altitude, humidity, etc.) different across test sites

Is it appropriate to combine data across these OT when estimating reliability?

Example: UAS IOT&E Reliability

Metric (Aborts)	Test Event	Hours	Aborts	Value (hours) [80% CI]	Requirement	Comparison[1] of Reliability Data with 29 Palms
$MFHBA_{System}$	29 Palms	188.3	12	15.7 [10.6 – 24.1]	50 hours (≡82% probability of completing 10 hour mission)	
	Lejeune	20.9	5	4.2 [2.3 – 8.6]		p-value = 0.02
	USS Anchorage	24.4	2	12.2 [4.6 – 45.9]		p-value = 0.67✓
	All 3 Phases	233.6	19	12.3 [9.0 -17.1]		
	29 Palms & Anchorage	212.7	14	15.2 [10.6 – 22.5]		
$MTBA_{Surface}$ Components	29 Palms	379.6	6	63.3 [36.0 – 120.4]	240 hours	p-value = 0.66✓
	Lejeune	90.6	2	45.3 [17.0 – 170.4]		
	USS Anchorage	72.9	2	36.5 [13.7 – 137.1]		N/A
	29 Palms & Lejeune[2]	470.2	8	58.8 [36.2 – 101.0]		
$MFHBA_{Air}$ Vehicle	29 Palms	188.3	6	31.4 [17.9 – 59.7]	60 hours	
	Lejeune	20.9	3	7.0 [3.1 – 19.0]		p-value = 0.053
	USS Anchorage	24.4	0	15.2 LCB		p-value = 1 ✓
	All 3 Phases	233.6	9	25.9 [16.4 – 43.0]		
	29 Palms & Anchorage	212.7	6	35.5 [20.2 – 67.5]		

Note 1: Gray-Lewis Two Sided Test for Exponential Means
Note 2: Only 29 Palms and Lejeune data can be combined. The surface components differ for shipboard configuration.

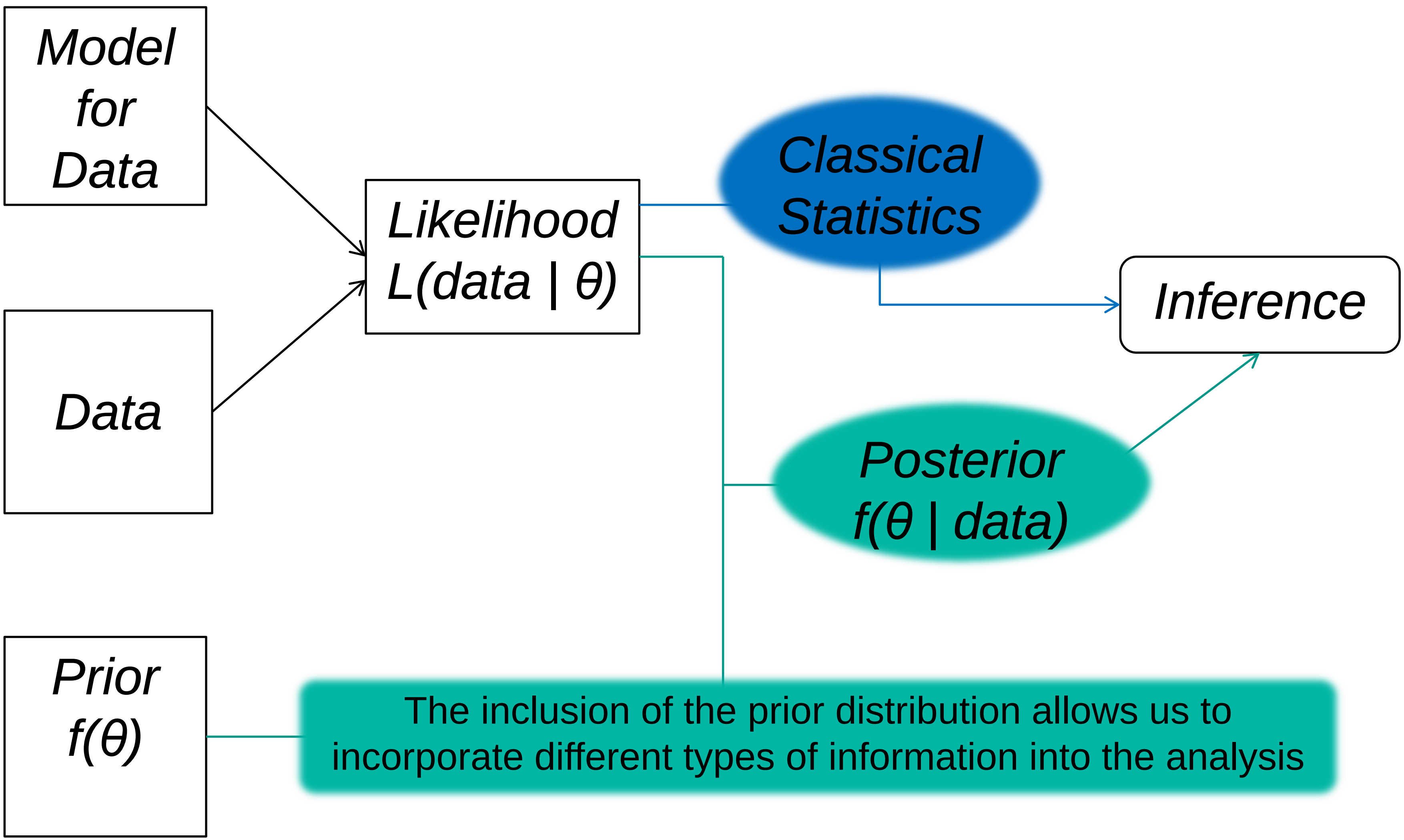
MFHBA – Mean Flight Hours Between Aborts
MTBA – Mean Time Between Aborts

15.2 $MFHBA_{System}$ ≡ 51.8% probability of completing 10 hour mission

Leveraging DT Data for an OT Assessment

- **Typically, we focus our reliability analysis to a single test period.**
  - But, shorter test periods, high reliability requirements, and/or few observed failures can result in little confidence in the reliability estimates.
- **We can employ statistical approaches to capitalize on all available data from multiple test periods.**
  - In support of DOT&E, IDA has begun to explore improved techniques for estimating reliability using data from multiple test periods
- **More advanced methodologies for assessing reliability are required.**
  - Censored Data
  - Generalized Linear Models
  - Bayesian Methodologies

Bayesian Methodology – The Big Picture



Example: Family of Armored Vehicles Reliability

- **Family of Combat vehicles*:**
  - Infantry Carrier Vehicle (ICV)
  - Antitank Guided Missile Vehicle (ATGMV)
  - Commander's Vehicle (CV)
  - Engineer Squad Vehicle (ESV)
  - Fire Support Vehicle (FSV)
  - Medical Evacuation Vehicle (MEV)
  - Mortar Carrier Vehicle (MCV)
  - Reconnaissance Vehicle (RV)

Vehicles share a high degree of commonality.

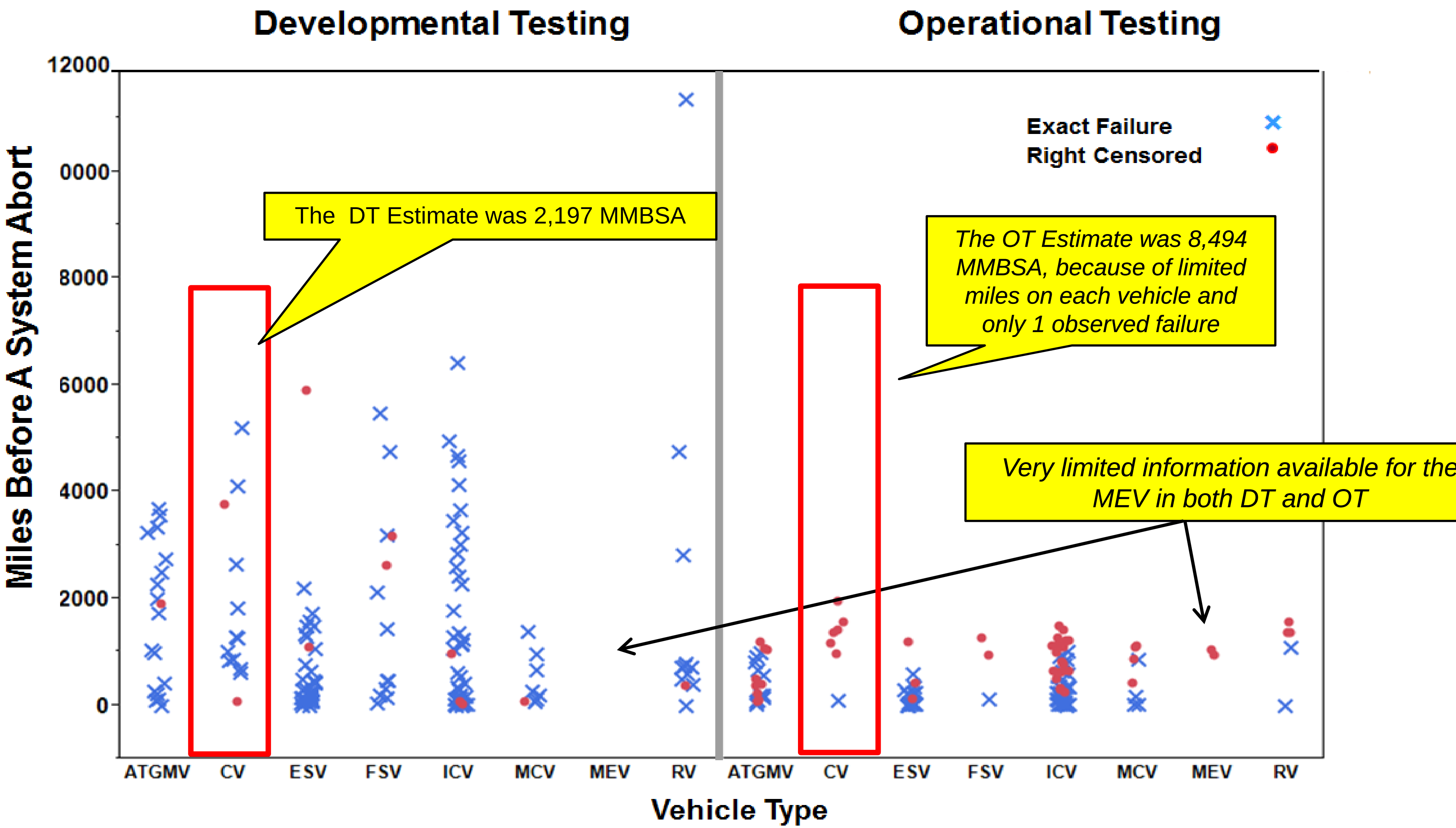
- **Reliability Requirements:**

  "The Armored Vehicle will have a reliability of 1000 mean miles between critical failure (i.e. system abort)"

- **Leveraging Information across two test phases: DT and OT**
  - There are known differences between DT and OT that should result in practical differences in their reliability estimates: Test length, Road conditions, vehicle drivers and mission durations varied between DT and OT.
  - Rather than consider each test phase and vehicle independently of each other, we can improve on the reliability analysis by using a statistical model to formally combine the data and make inference.

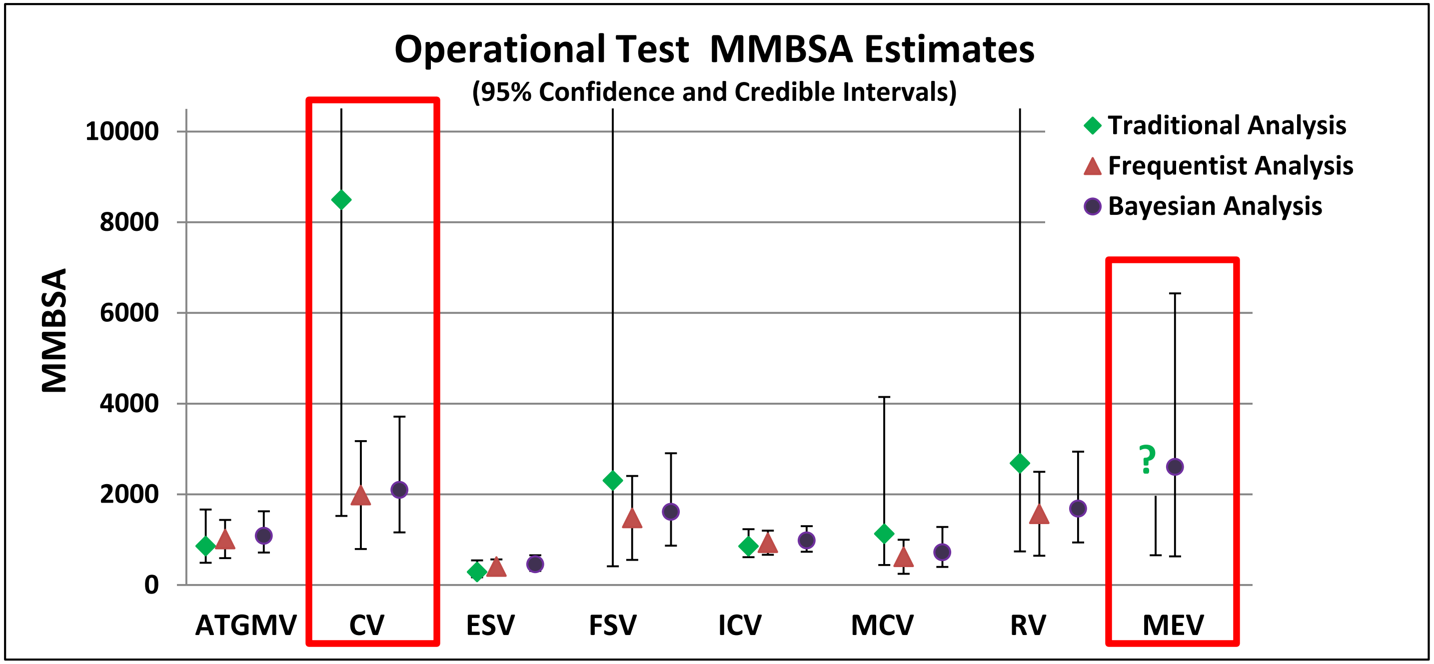
**The NBC RV was excluded from the study because of its different acquisition timeline.*

Family of Armored Vehicles



** A right censored observations occurs when the testing of the vehicle was terminated before a failure (i.e. system abort) was observed*

Family of Armored Vehicles



Traditional Analysis:

- Extremely wide confidence intervals!

Frequentist Analysis (Exponential Regression) & Bayesian Analysis:

- Mean Miles Between System Aborts (MMBSA) estimate and intervals calculated using DT and OT data
- Allows for a degradation in MMBSA from DT to OT
- Leverages all information
  - Better estimates of MMBSA and Tighter confidence intervals
- Bayesian Analysis allows for an estimate of the MEV MMBSA

Example: Ambulance Reliability

- **The primary mission of the Ambulance-equipped unit is medical evacuation.**

- **Limited User Test**
  - The Army conducted a Limited User Test (LUT) of the Ambulance to assess its capability to support rapid collection, evacuation, and pre-hospital life support of combat casualties. The LUT provided human factors, safety, and **reliability**, availability, and maintainability (RAM) data.

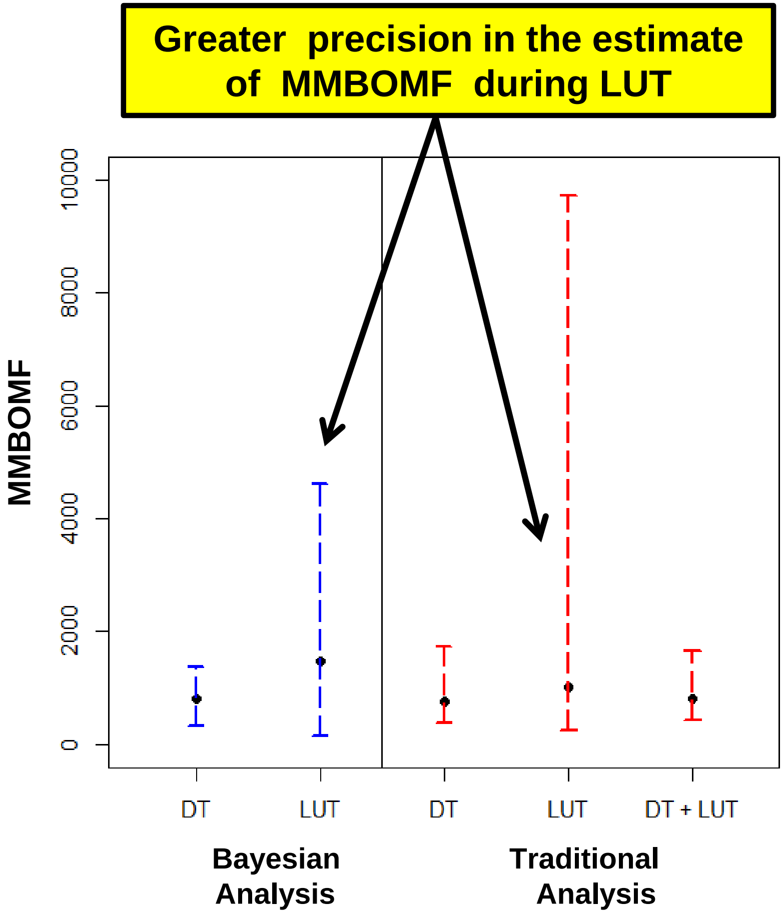
- **Reliability Requirements:**
  - These vehicles have a mean miles between operational mission failure requirement of at least 600 miles.

- **Leveraging information across two test phases: DT and LUT**
  - There are known differences between DT and LUT that should result in practical differences in their reliability estimates: test length, road conditions, vehicle drivers and mission durations varied between DT and LUT.
  - Rather than consider each test phase independently of each other, we can improve on the reliability analysis by using a statistical model to formally combine the data and make inference.

Ambulance: Reliability

- **These vehicles have a MMBOMF requirement of at least 600 miles.**
- **There was one OMF in LUT (1,025 miles) and four OMFs in DT (3,026 miles)**
  - One flat tire in LUT
  - Three flat tires and one air conditioner failure in DT
- **The Ambulance MMBOMF was estimated to be 1,479 miles during LUT and 824 miles during DT**
  - Point estimate and credible intervals calculated using DT and LUT data (Bayesian statistical method)

Method	Phase	MMBOMF	80% Confidence Interval
Bayesian Analysis	DT	824.4	(320.5, 1362.9)
	LUT	**1478.7**	**(141.4, 4610.8)**
Traditional Analysis	DT	605.2	(326.3, 1243.9)
	LUT	**1025**	**(263.5, 9758.5)**



Combining Data Requires Forethought

- **If the program wants to use DT data for OT:**
  - Data collection procedures need to be consistent with OT procedures
    - » Time between failures
    - » Failure modes identified
  - PM should note which failure modes (and which corresponding failures observed in testing) are addressed by corrective actions between test events

- **If the program wants to use data from earlier OT events for Initial or Follow-on Operational Test evaluation:**
  - Data collection procedures need to be consistent between OT events

What deviations from operational testing standards are acceptable and what deviations will preclude data from earlier test events from being used in evaluation?

Combining Reliabilities: Assessing the Reliability of a Complex System

- **One of the more difficult aspects of system reliability assessment is integrating multiple sources of information, including component, subsystem, and full system data, as well as previous test data or subject matter expert opinion.**
- **Reliability requirements for ships are often broken down into threshold for the critical or mission-essential subsystems.**
- **For example, the Capability Development Document for Small Shallow Ship (SSS) provides a reliability requirements for four functional areas.**
  - Sea Frame Operations, Core Mission, Mission Package Support, Phase II SUW Mission Package
  - The target reliability for Core Mission is 0.80 in 720 hours.
- **How do we assess the reliabilities of a system composed of multiple subsystems or components?**
  - Different Types of Data
    - » On-demand, continuous underway, continuous full
  - Not all subsystems have failures

The Bayesian approach to combining information from various subsystems/components and other sources to estimate full system reliability has many advantages.

USS Small Ship

Example: Reliability of a Multi-Mission Ship

- **Example: The Capability Development Document for SSS provides a reliability threshold for <u>Core Mission</u> functional area.**
  - The target reliability for Core Mission is 0.80 in 720 hours.

Test Data		
Critical Subsystem	Total System Operating Time	Operational Mission Failures
Total Ship Computing Environment (full-time)	4500 hours	1
Sea Sensors and Controls (underway)	2000 hours	3
Communications (full-time)	4500 hours	0
Sea Engagement Weapons (on-demand)	11 missions	2

➢ Assume the functional area is a series system: system is up if all subsystems are up.

Data are notional.

Comparison of Results

	Classical MTBOMF	Classical Reliability at 720hrs	Bayesian MTBOMF	Bayesian Reliability at 720hrs
TSCE	4500 hrs (1156 hrs, 42710 hrs)	0.85 (0.54,0.98)	3630 hrs (1179 hrs, 6753 hrs)	0.73 (0.54,0.90)
SSC	667 hrs (299 hrs, 1814 hrs)	0.33 (0.09,0.67)	697 hrs (332 hrs, 1172 hrs)	0.31 (0.11,0.54)
Comm	> 2796 hrs*	> 0.77*	10320 hrs (1721 hrs, 18210 hrs)	0.83 (0.66,0.96)
SEW		0.82 (0.58,0.95)		0.77 (0.62,0.91)
Core Mission		?????		0.15 (0.05, 0.27)

Comm – Communications
MTBOMF – Mean Time Between Operational Mission Failures
SEW – Sea Engagement Weapons
SSC – Sea Sensors and Controls
TSCE – Total Ship Computing Environment

** A conservative 80 percent lower confidence bound; frequentist MTBF does not exist*

Qualitative Assessments of Reliability

- **Mission impact of reliability**
  - Reliability failures preclude mission accomplishment
  - Excessive failures cause low availability
  - Maintainers unable to keep up with pace of system failures if system operated at OMS/MP-level tempo
- **Investigation of failure modes**
  - Are particular failure modes driving reliability estimates?
  - Are particular subsystems more prone to fail?
  - Are failures based on system use or do parts arrive broken “out of the box”?
- **Impact of sparing & redundancy on reliability**
  - Redundancy may ameliorate impact of failures
  - Are sufficient spares available to maintain operational tempo?
  - Was the number of spares available to maintainers representative of real-world operations?
  - Field-level vs. depot-level maintenance
- **Do any observed failures modes have an impact on user safety?**
- **Are failures being charged to users or maintainers?**

Takeaway Points

- **Reporting Reliability**
  – Was the system sufficiently reliable to successfully conduct its mission?
    » What is the demonstrated reliability?
    » Did the system meet its requirement? If not, what is the operational impact?
    » How does the system's reliability compare to the legacy system?

- **Reliability Models**
  – To ensure estimates of reliability are accurate, choosing the correct statistical model is crucial.

- **Combining Information**
  – There are sound statistical approaches that can be used to capitalize on all available data in assessing the reliability of a system.