2013 Reliability Survey

Matthew Avery
Daniel Butler
Douglas Peek
Project leader: Laura Freeman

IDA

Outline

- **Purpose, Changes, and Background**
- **Survey Population**
- **Policy Implementation & Impact on Reliability Outcomes**
- **Adequacy of Reliability Test Program**
- **Summary of Results**

Purpose

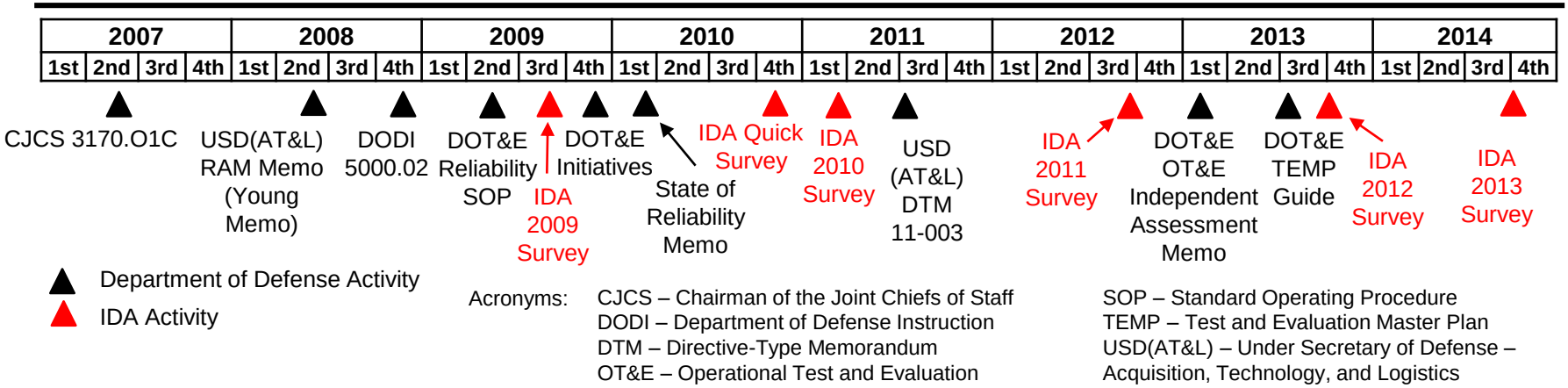
In this presentation, we assess responses from IDA research staff to a survey focused on the reliability efforts of programs that submitted either a Test and Evaluation Strategy (TES) or a Test and Evaluation Master Plan (TEMP) to DOT&E or had an Operational Test (OT) during FY13.

- **The objectives of this review are to:**
  - *Understand the extent to which programs under oversight are implementing reliability-focused policy guidance from the Department of Defense.*
  - *Assess whether implementation of this guidance is leading to improved reliability.*
  - *Determine whether oversight programs have a coherent strategy for testing reliability goals in OT.*

- **This year's survey addresses the following topics:**
  - Does the program include a reliability growth program (plan)?
  - Does the program evaluate reliability maturation throughout all program phases (track)?
  - Does the program evaluate reliability growth potential during development (predict)?
  - Did the program meet reliability goals in operational testing?
  - Did the planning process constitute a coherent reliability test program?

Acronyms on this slide: DOT&E – Director, Operational Test and Evaluation

Background

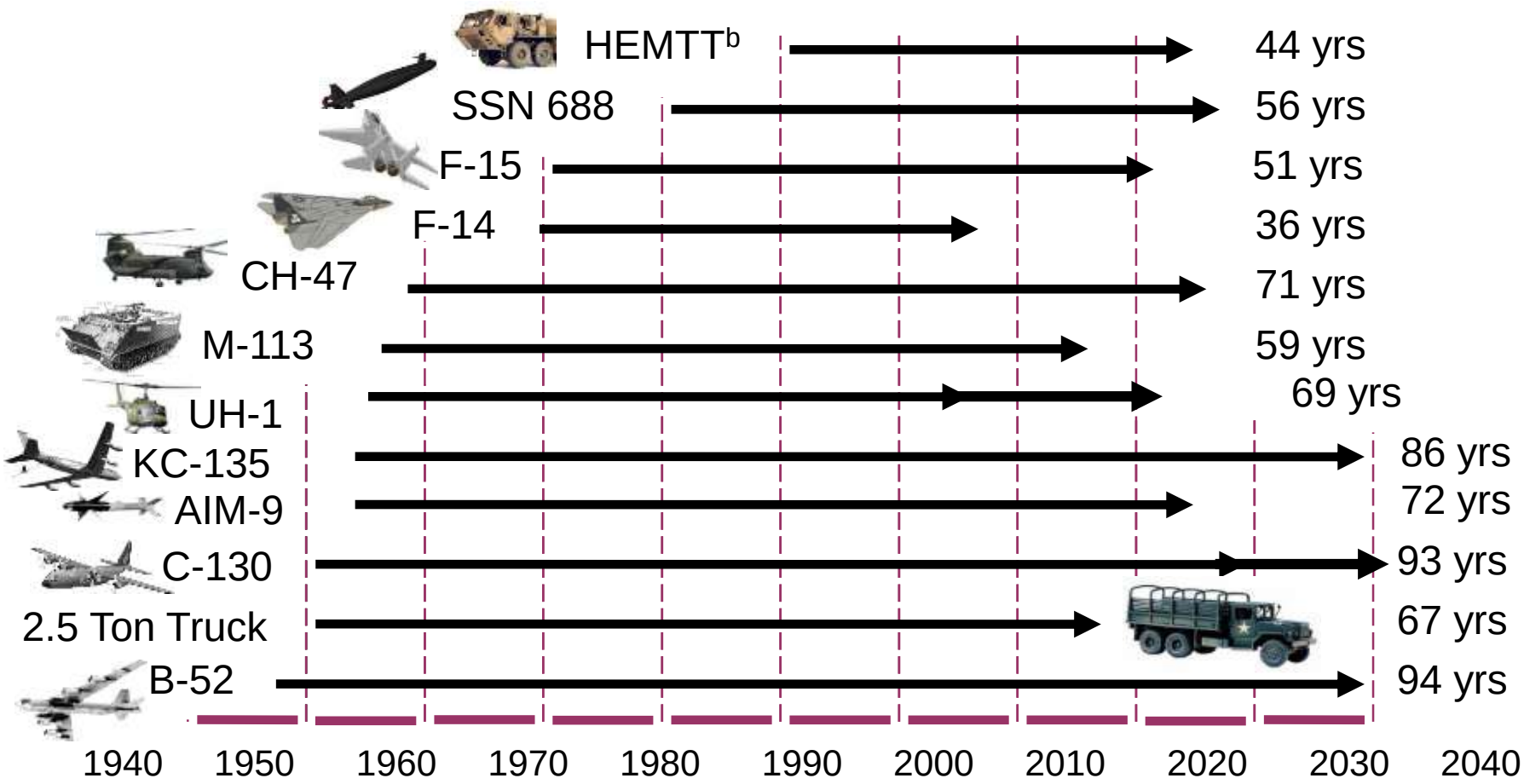


- **Improving reliability for the Department of Defense (DoD) systems continues to be a priority for the DOD.**

- **Survey Overview**
  - The survey included questions from the 2013 Survey considering three overarching areas in assessing whether programs are completing the activities necessary to develop reliable systems:
    - » Reliability growth strategy
    - » Reliability growth tracking
    - » Reliability in operational testing.
  - New questions were added to specifically look at the OT planning process.
  - Survey responses from previous years provided a baseline for comparison with 2013 responses.

Motivation for Reliability Growth Planning

- **Systems are in service for a long time, which drives up O&S costs[a]**

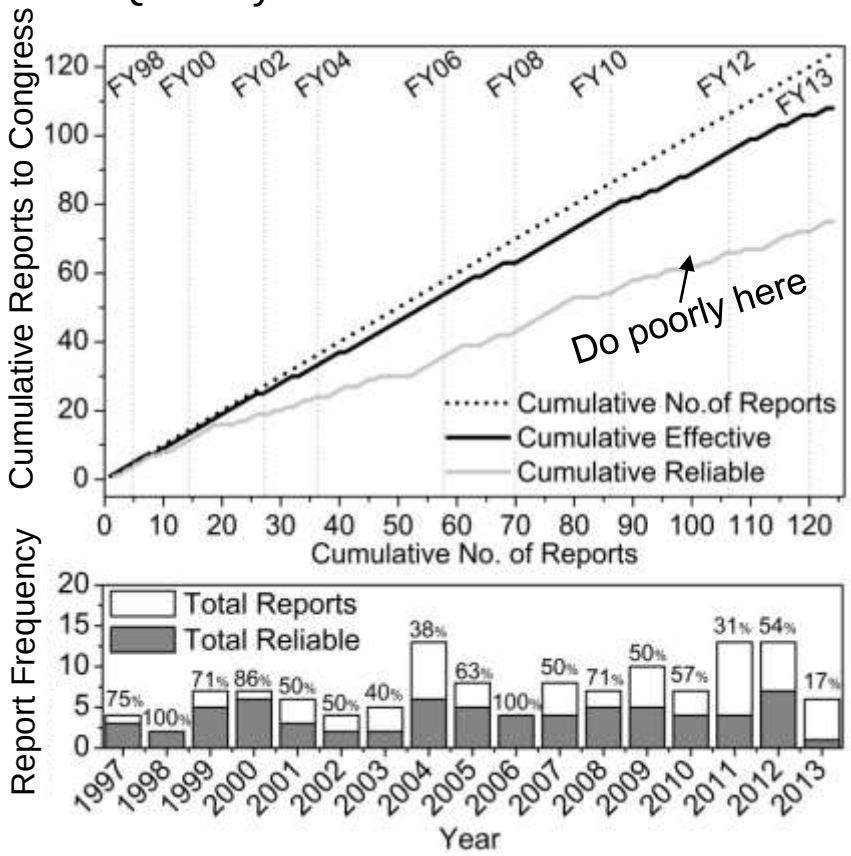


ORIGINAL SOURCE: John F. Phillips DUSD (L)

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

Motivation for Reliability Growth Planning

- **Why do it?**
  - Improve system reliability/meet thresholds
  - Optimize test resources
  - Quantify risks
  - Reduce Operations and Sustainment (O&S) Costs
  - Set 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. AEC/AMSAA Reliability Course Notes," 21 Aug 2011.

Majority of cost here

DOT&E TEMP Guidebook

- **All programs should have a comprehensive reliability program**
  - The program should employ a reliability growth strategy until all requirements are satisfied
  - An overview of the program should be documented in the TEMP
  - Areas of interest:
    - » Reliability allocations & block diagrams/system architectures
    - » Failure definitions and scoring criteria (FDSC)
    - » Failure mode, effects and criticality analysis (FMECA)
    - » System loads and use profiles
    - » Dedicated test events for reliability
    - » Growth testing at the system and subsystem level
    - » Failure reporting analysis and corrective action system (FRACAS)

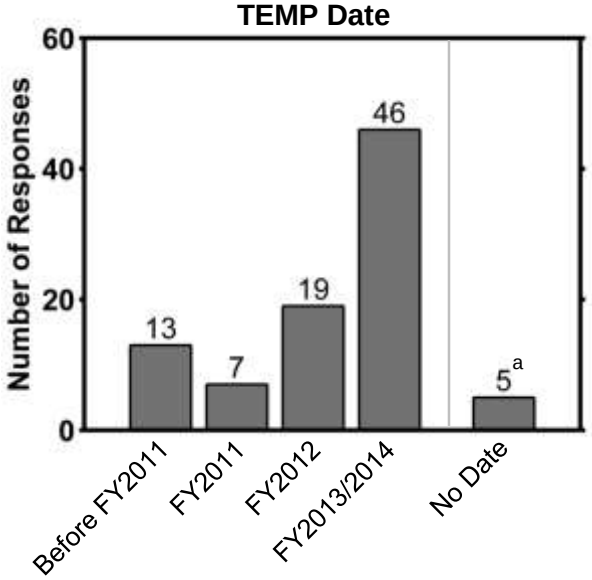
- **The reliability growth strategy should also be documented in the TEMP including:**
  - Initial reliability estimates
  - Reliability growth planning curves
  - Sufficient test time to discover failure modes
  - Sufficient time to implement fixes
  - Tracking curves
  - Entrance and exit criteria for each phase of test
  - Operating characteristic curves to assess risk

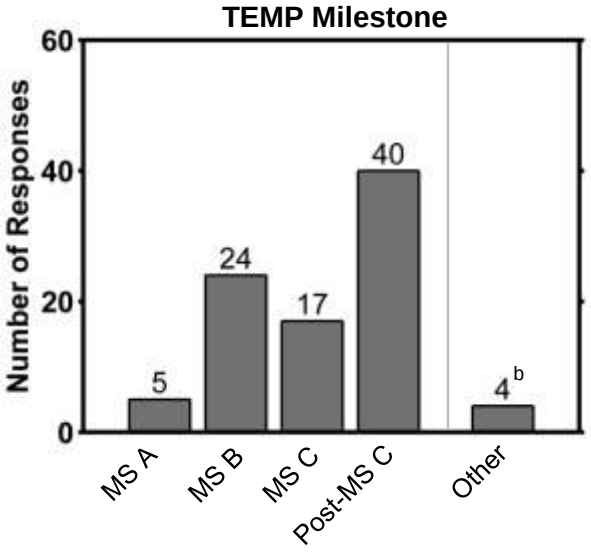
Outline

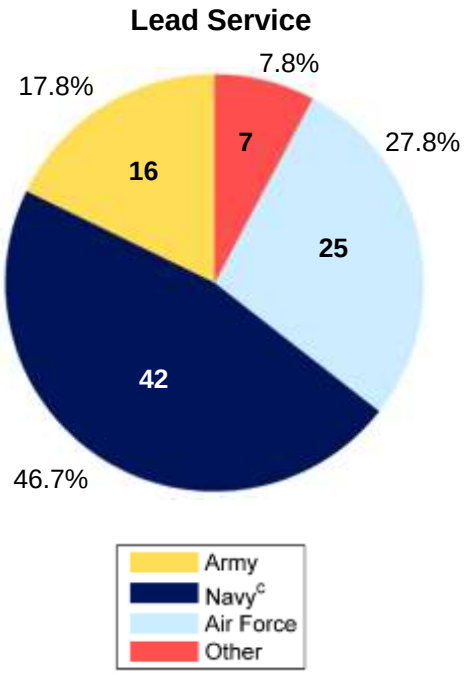
- Purpose, Changes, and Background
- **Survey Population**
- Policy Implementation & Impact on Reliability Outcomes
- Adequacy of Reliability Test Program
- Summary of Results

Survey Population
2013 Survey

- **The 2013 Survey focused on programs that submitted a TES/TEMP or had an Operational Test (OT) in Fiscal Year (FY) 2013; 90 responses:**
  - 59 of 90 programs had an OT in FY13/14.
  - 46 of 90 programs had a TES/TEMP approved in FY13/14.
- **44% (40/90) of responses were for post-MS C programs.**
- **As in previous surveys, Navy programs accounted for nearly half of all responses.**







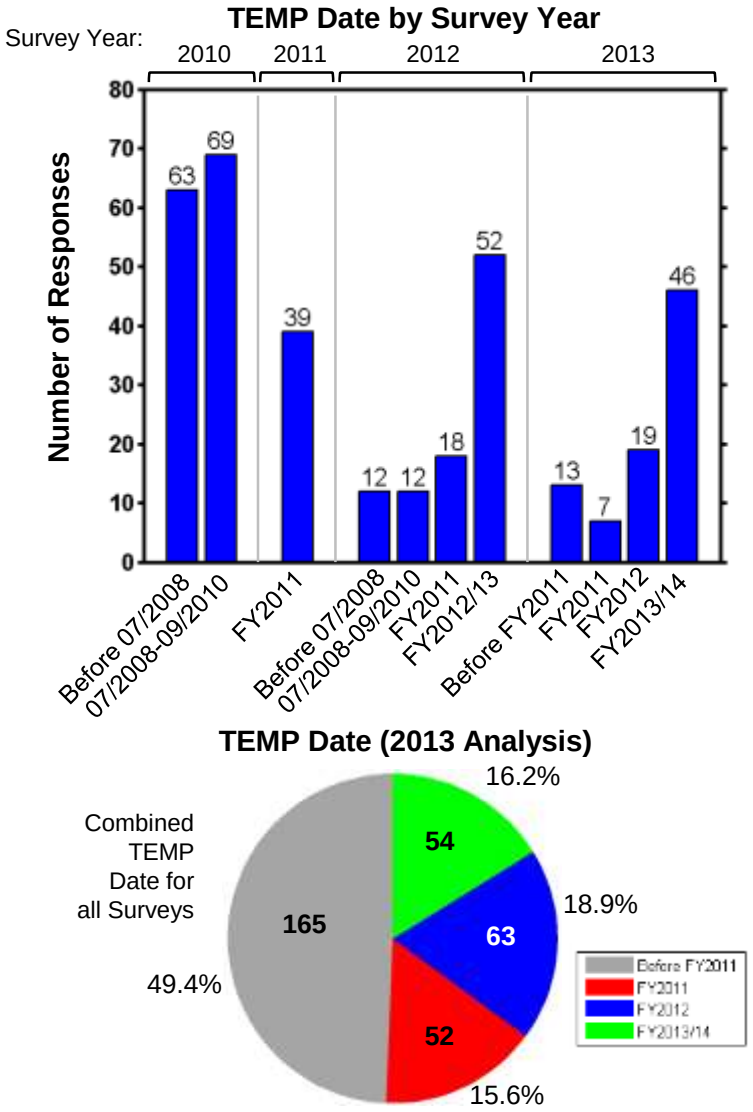
[a] 20 mm Fixed Forward Firing Weapons (FFFW) for the MH-60; AN/SQQ-89A(V)15 Surface Ship Undersea Warfare Combat System; Information Transport System Increment 2 (ITS2); Joint Tactical Radio System Handheld Manpack, Small Form Fit Manpack Radio; RQ-4B Block 40

[b] 20 mm Fixed Forward Firing Weapons (FFFW) for the MH-60; AN/SQQ-89A(V)15 Surface Ship Undersea Warfare Combat System; Joint Tactical Radio System Handheld Manpack, Small Form Fit Manpack Radio; RQ-4B Block 40

[c] Includes four Marine Corps Programs

Acronyms on this slide: TES – Test and Evaluation Strategy; TEMP – Test and Evaluation Master Plan; MS - Milestone

Survey Historical Analysis Approach



- **Data from past surveys was included to assess trends over time**
  - Top chart shows responses by survey year. (347 programs)
  - Some programs appeared in multiple surveys
  - Duplicate entries were removed from analysis unless there was a new TEMP with the most recent survey entry retained
  - Bottom chart shows total responses with duplicates over multiple survey years removed. (334 programs)
- **Survey years had different selection criteria**
  - 2010 attempted to include all oversight programs
  - 2011 only included programs that submitted a TEMP
  - 2012 and 2013 included programs that submitted TEMPs as well as programs that had an OT
- **Unlike previous years, all survey responses with TEMP dates prior to 2011 were combined in our analysis.**

Acronyms on this slide: TEMP – Test and Evaluation Master Plan

Outline

- Purpose, Changes, and Background
- Survey Population
- **Policy Implementation & Impact on Reliability Outcomes**
- Adequacy of Reliability Test Program
- Summary of Results

Policy implementation over time

- **The following have significantly improved since 2011:**
  - Proportion of programs having a reliability growth strategy
  - Proportion of programs including the strategy and growth curves in the TEMP
  - These positive trends have leveled off around 90% between 2012 and 2013

- **Other areas including the use of reliability growth curves, growth curves growing beyond the reliability threshold, and calculating growth potential, there is no statistical evidence of improvement.**

	Pre-FY2011	FY2011	FY2012	FY2013	p-value[1]
Have a reliability growth strategy	63% (90/143)	71% (36/51)	89% (55/62)	89% (48/54)	<0.001
Document reliability growth strategy in the TEMP	57% (51/90)	80% (28/35)	93% (51/55)	92% (44/48)	<0.001
Program has reliability growth curves	64% (55/86)	69% (24/35)	62% (34/55)	67% (32/48)	0.9118
Incorporate reliability growth curves into the TEMP	42% (23/55)	67% (16/24)	82% (28/34)	84% (27/32)	<0.001
Growth curves grow beyond requirement	75% (6/8)	67% (2/3)	50% (4/8)	72% (18/25)	0.7813
Programs have intermediate reliability growth goals	60% (6/10)	33% (2/6)	43% (6/14)	59% (23/39)	0.4505

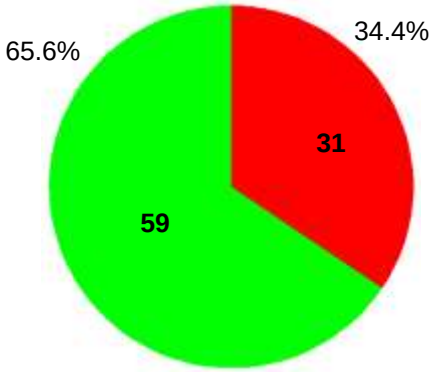
- **Inclusion of the reliability growth plan in the Statement of Work varies by service**

	Army	Navy	Air Force	Other	p-value[2]
Reliability growth plan included in Statement of Work	67% (10/15)	29% (10/35)	52% (11/21)	0% (0/5)	0.0121

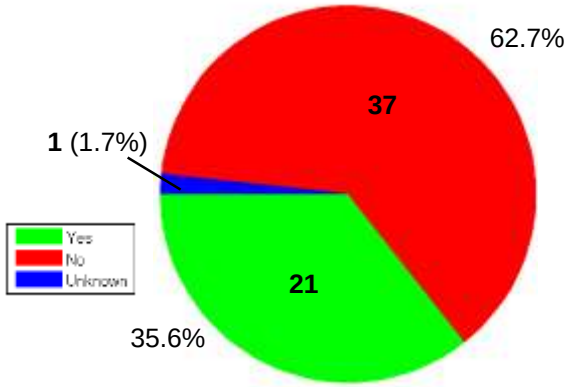
1. P-values based on ordinal regression. Small values show evidence for changes over time
2. P-value based on Pearson chi-square goodness of fit test. Small value shows differences between services.

Use of RAM-based Entrance Criteria for Operational Tests (OT) in 2013 Survey

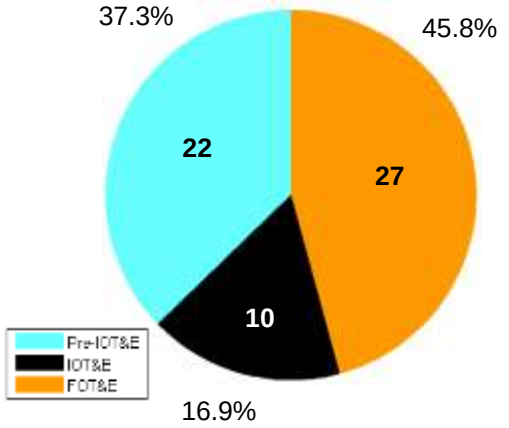
Did the Program have an OT (Q8)?

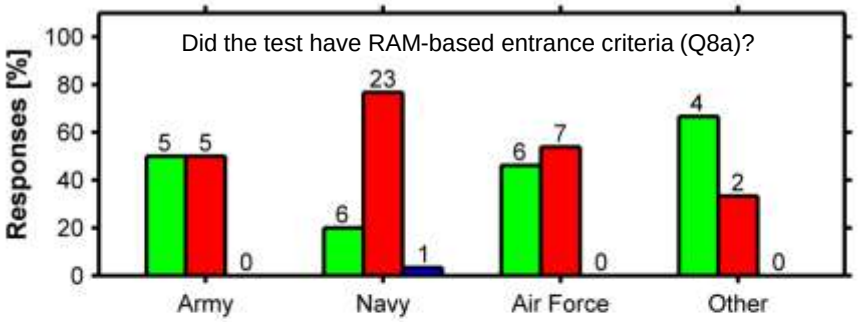


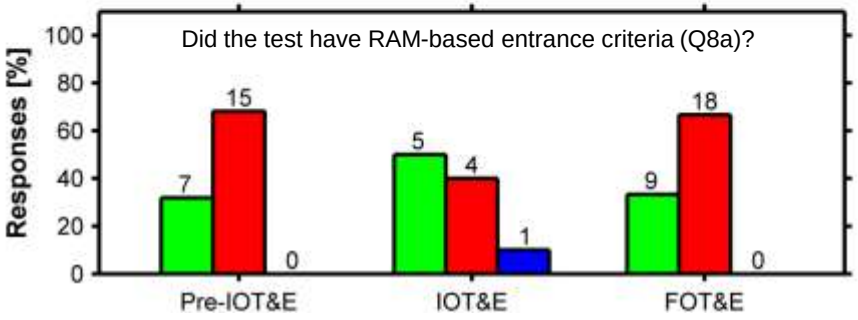
Did the test have RAM-based entrance criteria (Q8a)?



Type of OT





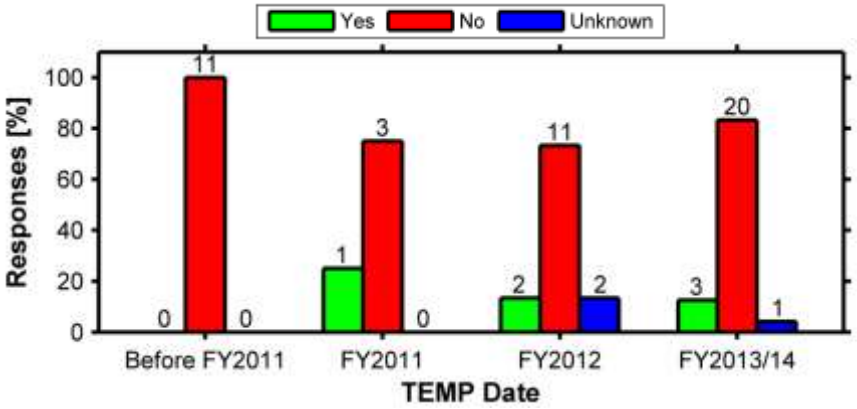


- **66% (59/90) of programs in the 2013 survey conducted an FY13/14 OT (Q8).**
  - The Navy had the largest number (51%) of OT events in FY13/14.
  - 17% (10/59) of OTs were IOT&E (compared to 41% in the 2012 survey).
- **36% (21/58) of programs in the 2013 survey with known responses had RAM-based entrance criteria for their OT (Q8a).**
  - 44% (4/9) with an IOT&E and 67% (18/27) with an FOT&E did not have RAM-based entrance criteria.
  - 79% (23/29) of Navy programs with known responses did not have RAM-based entrance criteria.

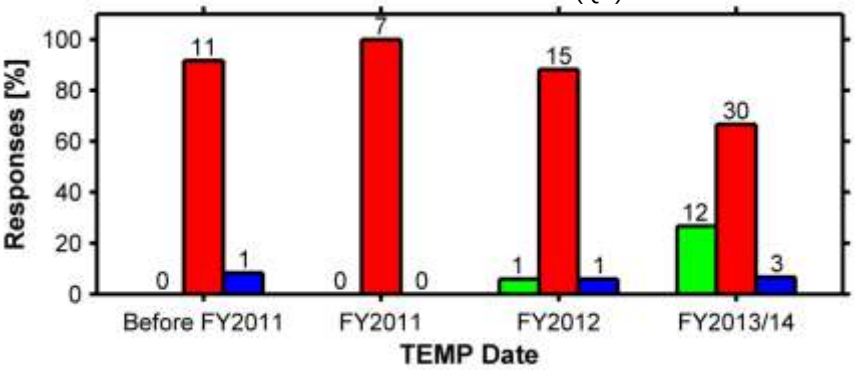
Acronyms on this slide: TEMP – Test and Evaluation Master Plan; RAM – Reliability, Availability, and Maintainability; IOT&E – Initial Operational Test and Evaluation; FOT&E – Follow-on Operational Test and Evaluation

Intermediate Goals and Planning OT using Power and Confidence

Did the system have an intermediate reliability goal for the FY13/14 OT (Q8b1)?



Does the TEMP/TES and/or test plan describe the producer (power) and consumer risk (confidence) for passing the reliability threshold at IOT&E/FOT&E (Q9)?



- **While 59% (23/40) of FY2013/14 programs had intermediate goals, most OTs do not have intermediate goals associated with them. (Q5d1/Q8b1)**
- **29% (12/42) programs with TEMPs approved in FY13/14 discuss power and confidence for passing their reliability threshold. (Q9)**
  - Among known responses, programs with new TEMPs were significantly (p-value=0.0312) more likely to discuss power and confidence
  - Many reliability policies were correlated with discussing power and confidence
    - » Growth curves in the TEMP (p=0.065)
    - » Reliability growth beyond threshold (p=0.0369)
    - » Documented intermediate goals (p=0.106)
    - » Intermediate goals based on demonstration of thresholds at OT events (p=0.0670)

Acronyms on this slide: TEMP – Test and Evaluation Master Plan

Which policies are associated with improved reliability outcomes?

- **RAM-based entrance criteria**
  - Meeting RAM-based entrance criteria was positively associated with demonstrating reliability at or above threshold at operational tests.
    - » Same result in 2012 survey
- **Inclusion of reliability growth strategy in Statement of Work (SoW)**
  - Programs that include their growth strategy in their contracting documents were more likely to demonstrate their thresholds at OT.
  - Trend is robust across all branches of military
- **Additional factor: Test phase**
  - Programs further along acquisition timeline are more likely to demonstrate their reliability threshold at OT events.

*Factors **positively** associated with demonstrating reliability threshold*		Did the system demonstrate reliability at or above threshold?	P-value[1]
Were RAM-based entrance criteria met?	Yes	67% (8/12)	0.0833
	No	0% (0/4)	
Does the SoW include the reliability growth strategy?	Yes	70% (7/10)	0.1134
	No/Unknown	33% (8/24)	
Test Phase	Pre-IOT	23% (3/13)	0.1512
	IOT	44% (4/9)	
	Post-IOT	57% (12/21)	

1. P-value based on Pearson chi-square goodness of fit test.

Outline

- Purpose, Changes, and Background
- Survey Population
- Policy Implementation & Impact on Reliability Outcomes
- **Adequacy of Reliability Test Program**
- Summary of Results

Assessing Reliability Test Program Adequacy

In previous years, we were unable to address the overall adequacy of reliability test programs based on survey results. To address this, we added three questions to this year's survey which asked for objective values.

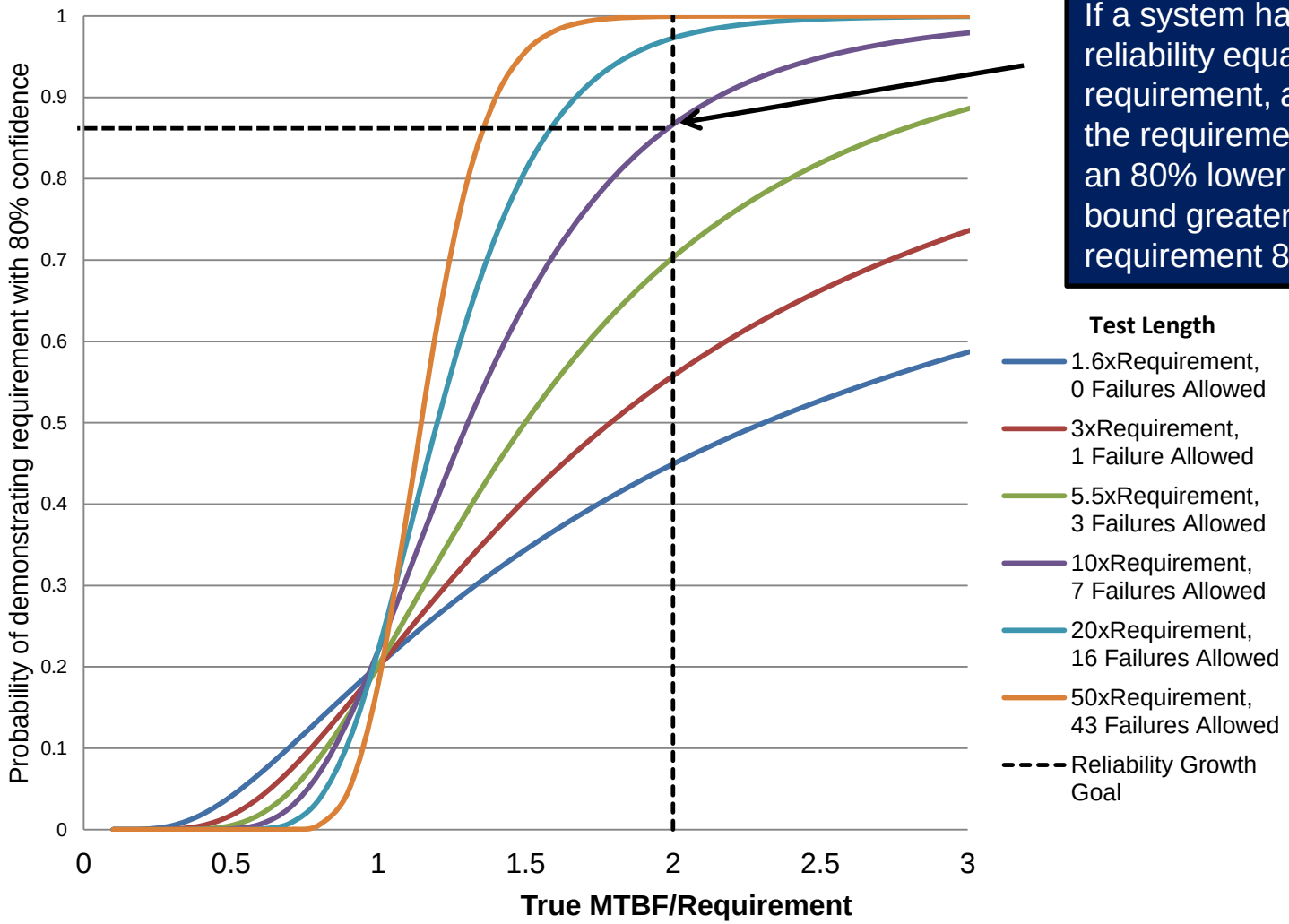
- **New questions for the 2013 survey:**
  - What is the IOT/FOT test length/test size described in the TEMP or test plan?
  - What is the system's reliability requirement?
  - What is the ultimate reliability growth goal?

- **Current DOT&E Guidance:**
  - IOT should provide sufficient data to assess system reliability with statistical confidence
    - » No guidance is given for level of statistical confidence

- **Using Operating Characteristic (OC) curves, the values asked about in the questions above (test length, requirement, and reliability goal) determine the statistical confidence a test is sized for.**

- **While the statistical properties of a test do not determine its adequacy, they provide an objective measure of how much we are learning about reliability based on operational testing.**

Example OC Curve (80% Confidence Level)



If a system has achieved reliability equal to 2x the requirement, a test lasting 10x the requirement will achieve an 80% lower confidence bound greater than the requirement 87% of the time.

Operating Characteristic Curves

- **OC curves are used to find power and confidence of a test.**
- **Requirement**
  - Value to demonstrate with statistical confidence
- **True failure rate (x-axis)**
  - This is the failure rate a system needs to achieve in order to demonstrate the requirement with confidence for the given test length.
  - In the PM2 model, this is also the reliability growth goal.
- **Growth factor**
  - Ratio of the true failure rate to the requirement
  - For subsequent analysis, we estimate this value by determining what true failure rate much be achieved for a test to meet certain power and confidence thresholds.
- **Test length (y-axis)**
  - Test must be long enough to demonstrate requirement.
- **For a test to have 80% power to demonstrate the requirement with 80% (or 50%) confidence, these numbers must "add up".**
  - If the requirement and true failure rate are close, the test must be longer.
  - If the test time is restricted, the system must grow to a larger true failure rate to demonstrate the requirement.
  - If the test time is restricted and the system's growth factor are limited, the test may not be very likely to demonstrate the requirement with the desired level of confidence.

Reliability Requirements for Programs Under DOTE Oversight

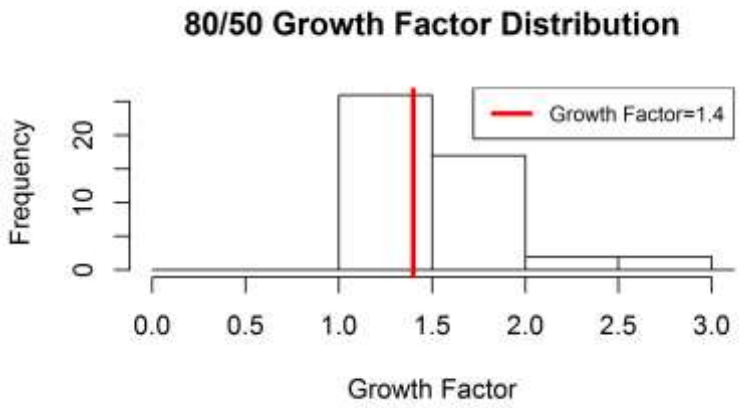
- **55 reliability requirements supported OC curve analysis**
  - Had both requirement and planned test
    - » Continuous (failure time) requirement
  - 44 unique programs (Some programs gave multiple requirements.)

- **Many programs did not have requirements appropriate for this type of analysis.**
  - 7 are based on system availability
  - 12 are binary failure/no failure requirements
  - 11 have no reliability requirement
  - 11 requirements were classified
    - » Further investigation of these showed results consistent with systems with unclassified requirements
  - 5 had no future reliability testing planned
  - 3 additional software-based systems

Do surveyed programs have reliability growth strategies that "add up"?

- **Using OC curves, we determined the true failure rate each program would need to achieve in order to meet both 80/80 & 80/50 criteria for the test length and requirement given in the survey.**
  - When information about growth goals was included in the survey, (30 of 55 requirements), we compared the growth goal to the true failure rates we determined.
    - » Follow-up questions asked about these programs when there were major differences
  - To demonstrating requirement with high power/confidence, programs had to have true failure rates far in excess of requirement
    - » Test sizes appeared reasonable for programs to demonstrate 50% confidence with 80% power rather than 80% confidence with 80% power

Growth beyond requirement needed to achieve high power/confidence at IOT			
Power/Confidence Level	Growth Factor		
	<1.4	< 2	< 3
80/50	21	42	47
80/80	9	22	33



For most programs, IOT insufficient to assess reliability even at 80/50

Approaches for Improving IOT Reliability Assessments

- **Increasing the size of all IOT/FOTs to address reliability with statistical confidence and power is not reasonable.**
  - Sizing IOT/FOTs to meet 80/80 with a growth factor of 1.4 would require increasing test lengths by an average of 592%.
- **Alternative approaches (many currently in use) should be considered**
- **Provide guidance on the use of pre-IOT reliability data in OT assessments**
  - Many programs do this already, but approaches differ, and there is no current guidance on best practices.
  - All test plans should address how much early testing will be incorporated in reliability assessments and discuss the methods by which such data will be combined with data from operational tests.
- **Encourage programs to use lower-level reliability measures (such as Essential Function Failures (EFFs) or Essential Maintenance Actions (EMAs)) in reliability assessments in addition to more common mission failure-based metrics.**
  - Incorporates a larger share of the failure modes
  - Can reduce subjectivity of scoring
  - Increase power and confidence without requiring additional testing

Outline

- Purpose, Changes, and Background
- Survey Population
- Policy Implementation & Impact on Reliability Outcomes
- Adequacy of Reliability Test Program
- **Summary of Results**

Summary of Results

- **Early action on reliability is a crucial component for programs that successfully demonstrate their reliability threshold in OT.**
  - Programs whose reliability growth strategies are discussed in the Statement of Work or other contracting documents are more likely to demonstrate reliability at or above the required value during OT.
    - » First year this question was asked
  - Programs that met RAM-based entrance criteria are more likely to demonstrate reliability at or above the required value during OT.
  - The only other identified predictor of success is how far along in the OT timeline a program is.

- **Lengths of IOT/FOTs relative to requirements indicate that programs are sizing IOT/FOTs towards demonstrating reliability requirements rather than demonstrating them with statistical power.**
  - Most RSMs indicated that sufficient data was collected to evaluate system reliability.
    - » Many programs used additional data from DT or from deployed versions of the system to make up the difference.
    - » Other programs had such poor reliability that a shorter test was sufficient for making a determination with statistical confidence.
    - » In some cases, testing was sufficient but requirements were not operationally meaningful

Acronyms on this slide:
IOT – Initial Operational Test
RSM – Research Staff Member
FOT – Follow-on Operational Test
RAM – Reliability, Availability, and Maintainability

Summary of Results (continued)

- **The following have significantly improved since 2011:**
  - Proportion of programs having a reliability growth strategy
  - Proportion of programs including the strategy and growth curves in the TEMP
  - These positive trends have leveled off around 90% between 2012 and 2013
- **In other areas, including the use of reliability growth curves, growth curves growing beyond the reliability threshold, and calculating growth potential, there is no statistical evidence of improvement.**
- **While programs continue to implement portions of DOT&E-directed reliability guidance, there is no evidence that these policies improve reliability outcomes**

Acronyms on this slide: DOT&E – Director, Operational Test & Evaluation

Elements of a Well-Run RG Program

Realistic Reliability Growth (RG) Curve

- Based on funding
- Realistic assumptions

Reliability Growth Model is the "tip of the iceberg"

- Component Design for Reliability
- Built-In-Test Demonstration

Dedicated Test Events for Reliability

- Operational Testing
- Accelerated Life Testing
- Logistics Demo
- Integration Testing

- System-level values achieved before fielding
- Contract Spec
- Interim thresholds
- Entrance/Exit criteria
- Appropriate DT metric

Adequate requirements

Reliability Analyses

- Failure Mode Effects and Criticality Analysis
- Level of Repair
- Reliability Predictions

- Funding and time allotted with commitment from the management

Corrective Actions

Data collection, reporting, and tracking

- Failure Definition Scoring Criteria
- Failure Reporting and Corrective Action System
- Failure Review Board
- Field Data
- Reliability, Maintainability, Availability Working Group

- Independent DT/OT data collection
- Scoring/assessment conferences
- Root cause analysis

Backup

Summary of Statistical Methods Employed

- **Pearson Chi-Square**
  - For contingency table analysis, the Pearon Chi-Square test was used. This test is used to determine if the distribution of one factor is dependent on another factor. (Or "factors" as the case may be.) For example, we tested whether the proportion of programs that met their reliability growth goals was different for programs that had had met their entrance criteria vs. programs that failed to meet their entrance criteria.

- **Ordinal Regression**
  - For analysis that compared ordinal (ordered) factors, such as TEMP Date or Test Phase, an ordinal regression was used. This test accounts for the fact that one of the factors has a specific ordering. TEMPs approved in 2012 were approved before 2013 TEMPS but after 2011 TEMPs. Unlike the Pearson Chi-Square (which treats them as different but unordered categories), this approach accounts for the ordering, allowing for a test with higher power.

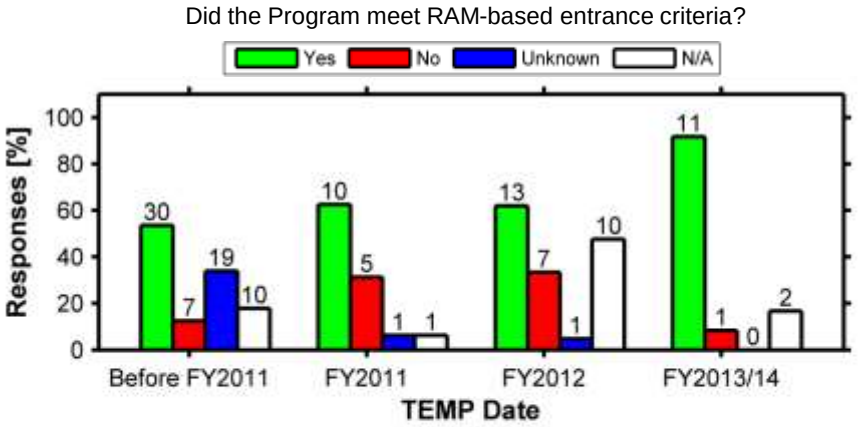
- **Cochran-Mantel-Haenszel (CMH) Test**
  - This test is used to assess the relationship between two categorical factors while adjusting for a third potentially confounding factor. A standard example is a clinical trial testing a new treatment against a placebo. In this example, the trial is taking place across *k* locations, all of which may have different standards of care. The CMH test determines whether there is a difference in patient outcome depending on whether patients were treated with the new drug or just given the placebo regardless of which location the test occurred at. This is analogous to the continuous response case of a one-way ANOVA with a random blocking factor.

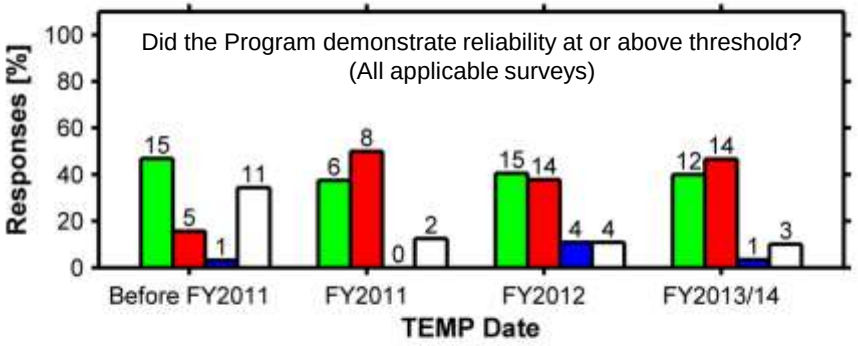
Statement of Work Impact on Reliability Outcomes by Service

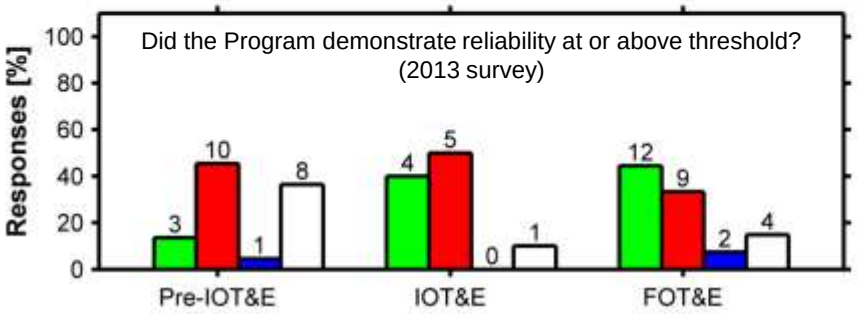
- **Across all service branches, programs that included their reliability growth strategy in the Statement of Work or other contracting documents were more likely to achieve their reliability requirement in OT**

		Did the system meet its requirement?				
		Service				P-value
		Army	Navy	Air Force	Other	
Does the system include its reliability growth plan in the SOW?	Yes	50% (2/4)	75% (3/4)	100% (2/2)	N/A (0/0)	0.0208
	No/Unknown	40% (2/5)	33% (4/12)	0% (0/3)	50% (2/4)	

Relationship Between Meeting RAM-based Entrance Criteria and Demonstrating Reliability in OT







- **92% (11/12) of applicable FY13/14 TEMP programs met their RAM-based entrance criteria.**
  - There were no significant differences by TEMP date category (p=0.3536 for known responses).
  - Air Force had the highest “Yes” response fraction.
  - Statistical analysis across all surveys suggests that having reliability growth curves is negatively related to achievement of RAM based entrance criteria (p=0.02284)
- **For programs with a known response, 44% (19/43) of 2013 survey programs that had an OT in FY13/14 demonstrated a reliability at or above the required value during OT.**
  - This was consistent with success rates from previous survey years
  - FOT&Es had the highest fraction of programs demonstrating reliability above the requirement.

Acronyms on this slide: TEMP – Test and Evaluation Master Plan

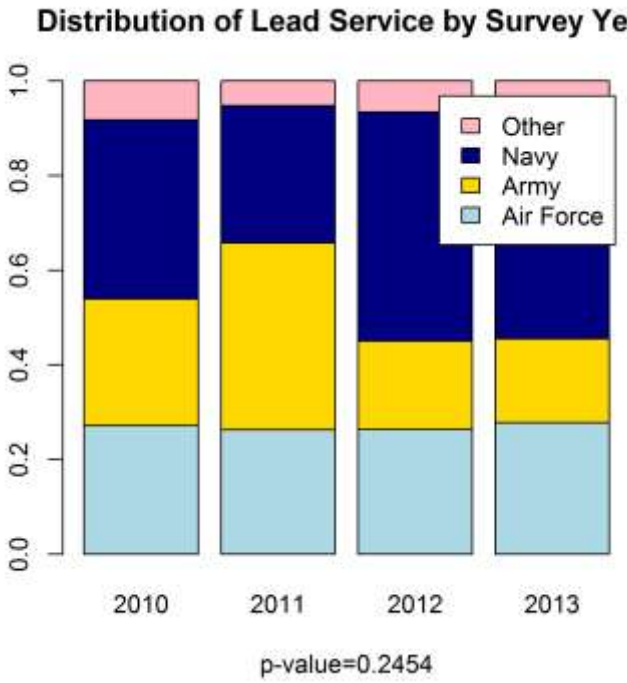
Which programs met their entrance criteria?

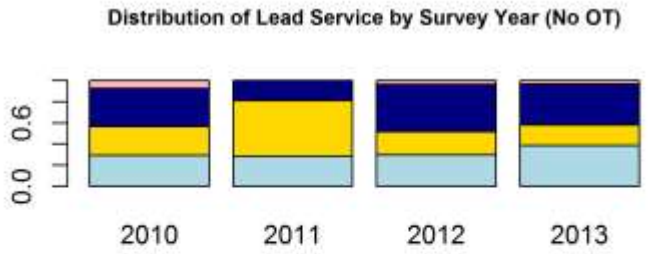
- **Many policies are weakly associated with meeting RAM-based entrance criteria**
  - 16 of 21 programs with RAM-based entrance criteria met these criteria (1 “Unknown” and 4 “No”)
  - Many other policies associated with robust reliability growth strategy positively associated with meeting RAM-based entrance criteria

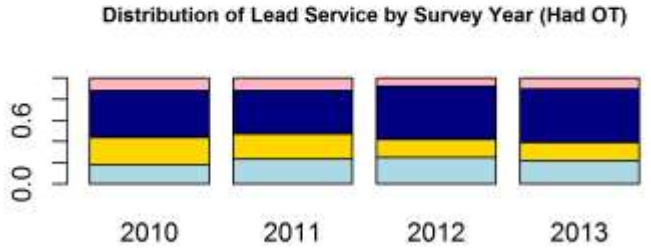
		Were RAM-based entrance criteria met?		P-value
		Yes	No	
Does the system have growth curves?	Yes	5	4	0.1457
	No	7	0	
Do the growth curves grow beyond the threshold?	Yes	2	4	0.2357
	No	3	0	
Given that there were intermediate goals, were these goals met?	Yes	2	0	NA
	No	0	2	
Does the TEMP discuss producer and consumer risk (power and confidence)?	Yes	2	2	0.2192
	No	12	2	
	No Req	2	0	

Distribution of Lead Service Over Time

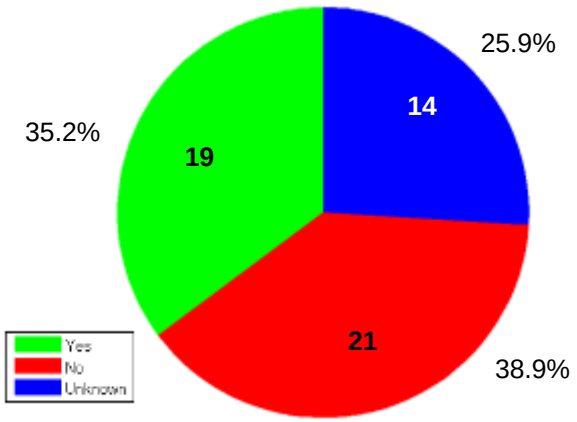
- **Over the length of the life of the survey, there has generally been more Navy programs than others.**
  - Especially pronounced in most recent years
  - 2011 is the exception, with many more Army programs
    - That survey year only included programs with OTs
  - P-value testing whether distribution has changed was marginal
  - Significant difference if we account for whether programs had OTs





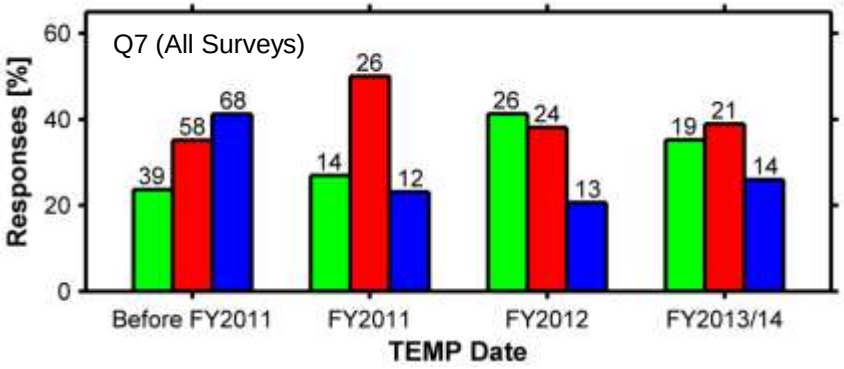


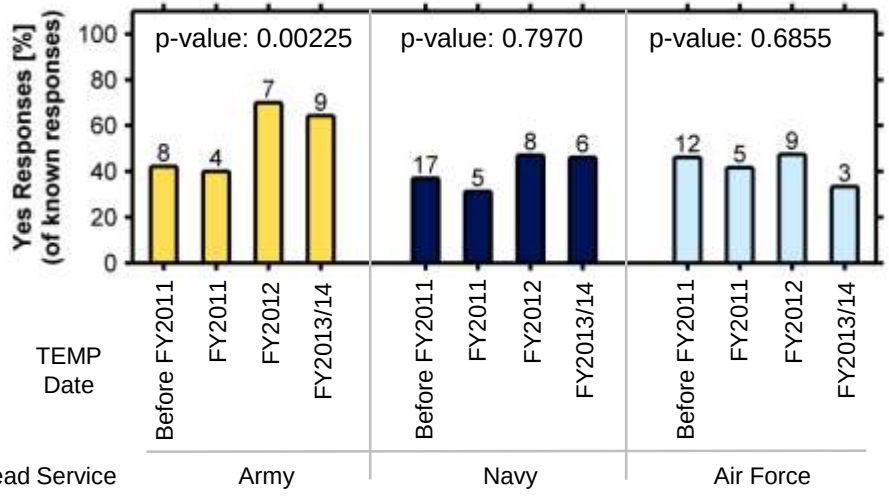
Do Programs Have a Process for Calculating the Reliability Growth Potential?



Q7 (TEMP Date: FY2013/14)

- **For programs with an FY13/14 TEMP and a known answer, approximately 50% (19/40) have a process for calculating reliability growth potential.**
- **Responses indicate fewer unknowns for later TEMP date programs compared to Pre-FY2011 TEMPs; otherwise, overall trends over time are insignificant.**
- **Army programs are most likely to calculate growth potential.**
  - Army is the only service to demonstrate improvement over time





Acronyms on this slide: TEMP – Test and Evaluation Master Plan

Number of Responses for Each Question

Question	Total[1]	2013	Total Known	2013 Survey Known	2013/14 TEMP Cat Known
5. Does the program have a reliability growth or improvement strategy?	334	90	310	88	54
5a. If the program does not have a reliability growth or improvement strategy, has the program met reliability requirements and completed testing?	81	13	81	13	6
5b. Does the TEMP describe the reliability growth or improvement strategy or reference where it can be found?	229	75	228	75	48
5c. Does the program have a reliability growth curve(s)?	229	75	224	75	48
5c1. Does the reliability growth curve(s) appear in the TEMP?	145	50	145	50	32
5c2. Does the growth curve(s) grow beyond the reliability threshold specified in the or derived from the requirements document?	48	48	44	44	25
5c3. What is the ultimate reliability growth goal? (e.g., the point on the reliability growth curve achieved at the IOT)	48	48	48	48	27
5d. Does the program have documented intermediate reliability growth goals?	117	75	112	71	39
5d1. Are the intermediate growth goals linked to the reliability growth curve(s)?	81	39	80	39	28
5d2. Are the intermediate growth goals based on demonstration of reliability thresholds at operational and/or developmental test events?	81	39	80	39	28
5e. Does the Statement of Work (or other contracting/specifications documents) include this reliability growth or improvement strategy?	73	73	42	42	23
6. Does the program routinely use reliability metrics to ensure reliability growth is on track to achieve requirements	334	90	269	85	51
6a. Please select all that apply	85	85	84	84	46
7. Has the program calculated its reliability growth potential?	334	90	227	69	40
8. Did the program have an operational test (DT/OT, IT, OA, LUT, IOT&E, FOT&E) in FY20XX?	334	90	334	90	54
8a. Were there RAM-based entrance criteria for this test event?	59	58	59	58	25
8a1. Were the RAM-based entrance criteria met prior to the test event?	142	26	124	24	14
8b. What type of operational test was it?	176	60	176	60	30
8b1. Did the system have an intermediate reliability goal for this test event?	54	54	51	51	23
8b2. Was the intermediate reliability goal met?	6	6	5	5	3
8c. Did the system demonstrate a reliability at or above the required value specified in (or derived from) the CDD/CPD during the operational test?	115	60	109	57	29
9. Does the TEMP/TES and/or test plan describe the producer (power) and consumer risk (confidence) for passing the reliability threshold at IOT&E/FOT&E?	81	81	76	76	42
9a. What is the IOT/FOT test length/test size described in the TEMP/TES or test plan? (e.g., “240 hours” or “2,000 miles” or “100 rounds”)	85	85	85	85	46
9b. What is the system's reliability requirement? (e.g., "10 MFHSA" or "43 MMBOMFF" or "90%")	85	85	85	85	46

1: Total for surveys that had a TEMP date. For example, there were 3 respondents to Question 5e that did not include a TEMP date.

Programs excluded from OC curve analysis

Requirement Category	Systems
Availability	Joint Space Operations Center (JSPOC) Mission System (JMS), Space-Based Infrared System Effectivity 5, Amphibious Assault Ship Replacement (LHA-6), Global Broadcast System (GBS), Public Key Infrastructure (PKI), C-5M Testing for Operational Flight Program (OFP), RQ-4B Block 40
Binary	B61 Life Extension Program (LEP) Tail Kit Assembly (TKA), Small Diameter Bomb Increment II (SDB II), Next Generation Diagnostic System, Guided Multiple Launch Rocket System with Alternative Warhead (GMLRS-AW), Miniature Air Launched Decoy Jammer (MALD-J) ADM-160C, Spider M7E1 XM1156 Precision Guidance Kit (PGK), Joint Warning and Reporting Network (JWARN), Rolling Airframe Missile (RAM), Joint Air to Surface Missile – Extended Range (JASSM-ER), Excalibur, Multi-static Active Coherent (MAC) Increment I Phase I on P-3 Aircraft
No Requirement	Information Transport System Increment 2 (ITS2), CVN 78 Aircraft Carrier (x5), Defense Readiness Reporting System (DRRS) Version 4.6, 20 mm Fixed Forward Firing Weapons (FFFW), F/A-18, Standard Missile-6 (SM-6), Cooperative Engagement Capability (CEC)
Classified Requirement	AN/BQQ-10 ARCI, AN/BYG-1 Combat Control System (Revision 7), AN/BLQ-10, OHIO-CLASS REPLACEMENT, Electronic Protection Improvement Program (EPIP), Mk 54 Mk 48 CBASS Torpedo, Mk 48 Heavyweight Torpedo with APB Spiral 4 Tactical Software
Test not yet planned/No future testing	Fleet Replenishment Oiler T-AO(X), CH-47F Cargo On/Off Loading System, Automatic Radar Periscope Detection and Discrimination (ARPDD)
Software	C2AOS/C2IS, B-2 EHF SATCOM, KMI
Other	Automatic Radar Periscope Detection and Discrimination (ARPDD),

Binary Reliability Requirements

System	Test Size	Requirement
Rolling Airframe Missile (RAM) Block 2	12	0.8
Multi-static Active Coherent (MAC) Increment I Phase I on P-3 Aircraft	781	0.8
Joint Air to Surface Missile – Extended Range (JASSM-ER)	21	0.85
Miniature Air Launched Decoy Jammer (MALD-J) ADM-160C	16	0.9
M982E1 Excalibur Increment 1b	32	0.9
B61 Life Extension Program (LEP) Tail Kit Assembly (TKA)	15	0.91
Small Diameter Bomb Increment II (SDB II)	55	0.91
Guided Multiple Launch Rocket System with Alternative Warhead (GMLRS-AW)	72	0.92
XM1156 Precision Guidance Kit (PGK)	48	0.92
Spider M7E1	28	0.96
Joint Warning and Reporting Network (JWARN)	40	0.98
Rolling Airframe Missile (RAM) Block 2	475	0.988

For most systems, tests were small relative to the number of data points necessary to achieve high confidence levels.

For example, for a requirement of 0.9, if the true system reliability was 0.95, 122 data points would be required to achieve 80/80.

Demonstrating noninferiority is somewhat easier. For the same 0.9 requirement, to demonstrate that the system is no worse than 10 percentage points worse (0.8) with 80/80 requires only 39 data points. (Based on nominal alpha value of 0.08, which has true false positive rate of 0.19).

Are any policies associated with worse reliability outcomes?

- **Reliability growth curves**
  - Programs with reliability growth curves were less likely to demonstrate reliability thresholds in operational tests.
  - When data from the FY2012 survey is included, results are still significant but the p-value increase (0.1727).
  - Trend disappears once service is taken into account.[2]

- **Reliability growth curves in the TEMP**
  - Programs with reliability growth curves in the TEMP were also less likely to demonstrate reliability thresholds in operational tests.

*Factors **negatively** associated with demonstrating reliability threshold*		Did the system demonstrate reliability at or above threshold?	P-value[1]
Does the system have growth curves?	Yes	7/22 (32%)	0.1109
	No	8/12 (67%)	
Do the growth curves appear in the TEMP?	Yes	2/16 (12%)	0.0021
	No	5/5 (100%)	

1. P-value based on Pearson chi-square goodness of fit test.
2. Based on Cochran-Mantel-Haenszel test. (p-value=0.4235)

Question 5: Growth Improvement Strategy/ Growth Curves

Question Hierarchy and Logic

5. Does the program have a reliability growth or improvement strategy?
(Yes, No, Unknown)

Yes → 5b. Does the TEMP describe the reliability growth or improvement strategy or reference where it can be found?
(Yes, No, Unknown)

5b. → 5c. Does the program have a reliability growth curve(s)?
(Yes, No, Unknown)

Yes → 5c1. Does the reliability growth curve(s) appear in the TEMP?
(Yes, No, Unknown)

5c1. → 5c2. 5c2. Do the growth curve(s) grow beyond the reliability threshold specified in the or derived from the requirements document?
(Yes, No, Unknown)

5c2. → 5c3. What is the ultimate reliability growth goal? (e.g., the highest point on the reliability growth curve)

5c. → No/Unknown → 5d. Does the program have documented intermediate reliability growth goals?
(Yes, No, Unknown)

5d. → Yes → 5d1. Are the intermediate growth goals linked to the reliability growth curve?
(Yes, No, No RGC, Unknown)

5d1. → 5d2. Are the intermediate growth goals based on demonstration of reliability thresholds at operational and/or developmental test events?
(Yes, No, Unknown)

5d. → No/Unknown → 5e. Does the Statement of Work (or other contracting/specification documents) include this reliability growth or improvement strategy?
(Yes, No, Unknown)

5. → No → 5a. If the program does not have a reliability growth or improvement strategy, has the program met reliability requirements and completed testing?
(Yes, No)

5. → Unknown → 5a.

5e. / 5a. → *Please provide any additional comments you might have on question series 5.*

Acronyms: FOT&E – Follow-On Operational Test and Evaluation IOT&E – Initial Operational Test and Evaluation MS - Milestone

Question 6: Ongoing Reliability Assessment

6. Does the program routinely perform assessments using reliability metrics to ensure reliability growth is on track to achieve requirements? (e.g., assessment conferences to assess fix effectiveness of corrective actions, use of reliability tracking models to assess progress, determination if the reliability is increasing with time, etc.)
(Yes, No, Unknown)

No/
Unknown

Please provide any additional comments you might have on question 6.

Yes

Proceed to question 7.

6. Please select all that apply:
-Assessment conferences to assess fix effectiveness
-Use of reliability tracking models to assess progress
-Determination if the reliability is increasing in time
-Defect Aging
-Failure Review and Corrective Action System (FRACAS)
-Plots of failure frequencies and/or MTBF over time
-Statistical comparison between growth planning curve and results (goodness of fit test)
-Projections updated based on DT/early test data
-Scoring conferences/JRMET
-Open Priority Tracking
-Other (please explain)

Acronyms: JRMET – Joint Reliability/Maintainability Evaluation Team

Question 7: Reliability Growth Potential

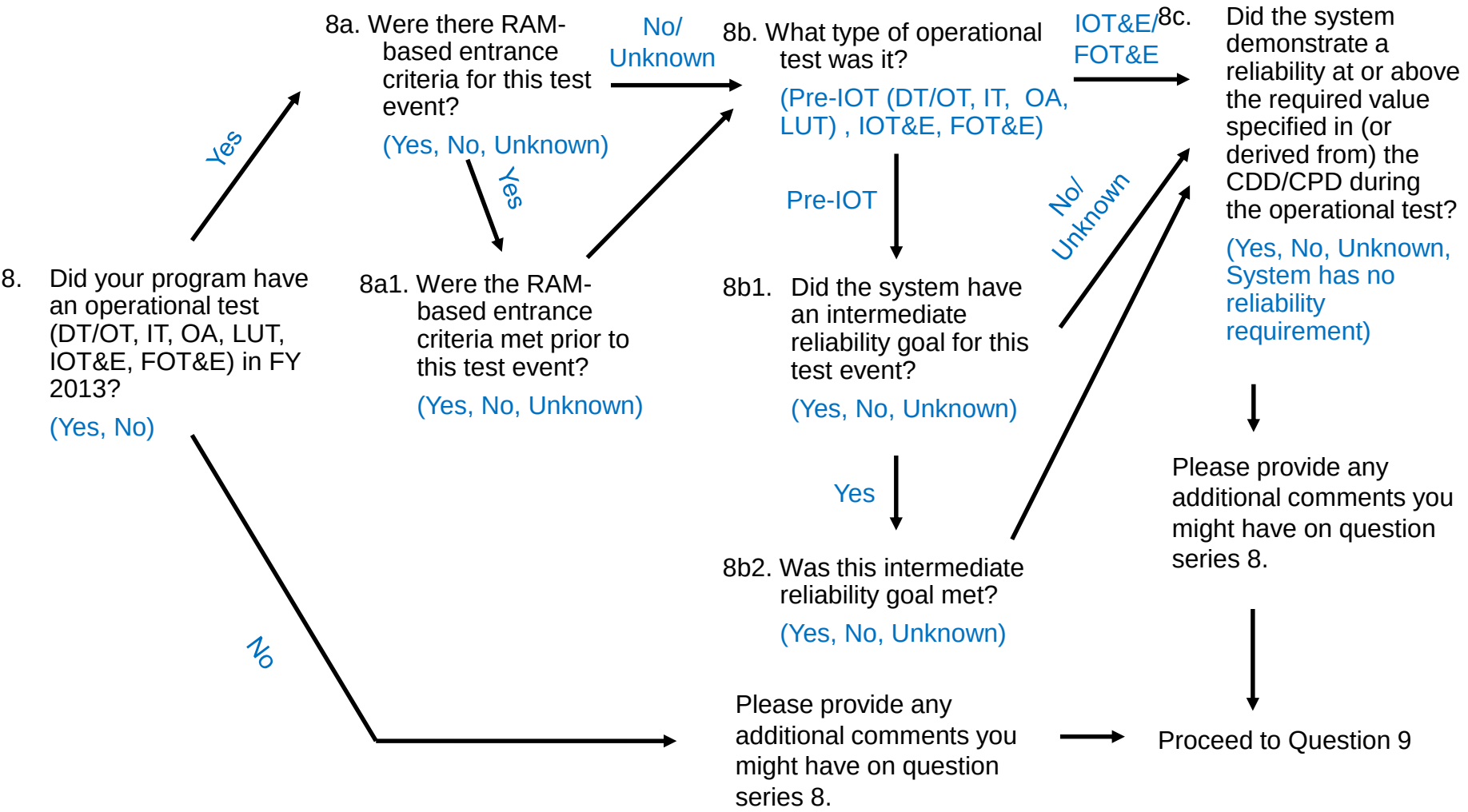
7. Has the program calculated its reliability growth potential?
   (Yes, No, Unknown)

→ Please provide any additional comments you might have on question 7.

→ Proceed to question 8.

Question 8: Entrance Criteria and Achieving Reliability Goals

Question Hierarchy and Logic



Acronyms:

DT – Developmental Test
OT – Operational Test
FOT&E – Follow-on Operational Test and Evaluation
OA/LUT – Operational Assessment/Limited User Test
IOT&E – Initial Operational Test and Evaluation
RAM – Reliability, Availability, Maintainability

Question 9: Adequate Test Size for Reliability

Question Hierarchy and Logic

9. Does the TEMP/TES and/or test plan describe the producer (power) and consumer risk (confidence) for passing the reliability threshold at IOT&E/FOT&E?

   (Yes, No, System has no reliability requirement)

9a. What is the IOT&E/FOT&E test length/test size described in the TEMP/TES or test plan? (e.g., “240 hours” or “2,000 miles” or “100 rounds”)

9b. What is the system’s reliability requirement? (e.g., “10 MFHSA” or “43 MMBOMF” or “90%”)

Please provide any additional comments you might have on question series 9.

End of Survey