REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
08-09-2008

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Test Operations Procedure (TOP)
1-1-030 RAM-D and ILS Analysis

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
RAM/ILS Engineering and Analysis Division (TEDT-AT-ADR)
US Army Aberdeen Test Center
400 Colleran Road
Aberdeen Proving Ground, MD 21005-5059

8. PERFORMING ORGANIZATION REPORT NUMBER
TOP 1-1-030

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Test Business Management Division (TEDT-TMB)
US Army Developmental Test Command
314 Longs Corner Road
Aberdeen Proving Ground, MD 21005-5055

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Same as item 8

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
Defense Technical Information Center (DTIC), AD No.: ADA486429

14. ABSTRACT
This TOP establishes guidelines to use when performing a RAM-D and/or ILS analysis during developmental testing. This document is intended for use in assessing RAM-D and ILS characteristics of various types of materiel, including individual items and systems.

15. SUBJECT TERMS

Availability	Maintainability
Durability	Logistic Support
Endurance	Reliability
Integrated Logistic Supportability	System Support and Sustainment

16. SECURITY CLASSIFICATION OF: a. REPORT	B. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	SAR	60	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

US ARMY DEVELOPMENTAL TEST COMMAND
TEST OPERATIONS PROCEDURE

Test Operations Procedure 1-1-030 8 September 2008
DTIC AD No.: ADA486429

RAM-D AND ILS ANALYSIS

1. SCOPE.

This Test Operations Procedure (TOP) establishes guidelines to use when performing a reliability, availability, maintainability, durability (RAM-D), or integrated logistic supportability (ILS) analysis during developmental testing. This document is intended for use in assessing RAM-D and ILS characteristics of various types of materiel, including individual items and systems. The great variety and complexity of this materiel makes it impractical to prepare procedures precisely matched to each system type or category. For this reason, material as presented herein is general; test personnel should prepare procedures necessary for the specific item being considered. These analytical procedures apply to all reliability and durability developmental tests which are conducted by ATEC Developmental Test Centers. These include both government and contractor testing in support of research, development, and acquisition programs.

Background:

The Army program for materiel readiness emphasizes the complementary attributes of reliability and maintainability. Reliability can be expressed as the probability that materiel will perform its intended function for a specified interval, i.e., remain in a state of readiness, without maintenance or with only planned maintenance. Reliability translates to usable and effective materiel in theater. Maintainability can be expressed as the efficiency, in time, tools, spare parts, and skills, with which materiel can be restored to readiness when maintenance is required. When assessing the reliability and maintainability of materiel, failure modes from varying types of performance tests are useful and can be examined in light of product improvement. However, valid meaningful RAM-D and ILS data can only be extracted from tests which are specifically designed to provide a sufficient quantity of operating time (or miles, or rounds, etc.) under conditions similar to those prescribed in the operational mode summary and mission profile (OMS/MP) of the system in question.

Endurance and durability tests are the typical tests during which RAM-D and ILS characteristics are evaluated:

a. An endurance test is a test which involves extended operation of one or more test items under cycles designed to simulate, under proving ground conditions, extended field use. The endurance test is the principal means of producing data for reliability and availability during development tests, and also is a major source of information on maintainability and human factors.

b. A durability test relates to the mathematical probability that an item or a major component thereof will be able to operate under defined conditions for a specified time or distance before requiring major overhaul, replacement, or salvage. A complete durability evaluation is seldom performed on development vehicles since large sample size, extensive test mileage, and considerable funding are required.

RAM-D and ILS analysis performed during developmental testing aids in the detection of significant problem areas in Army materiel early during the acquisition process, which allows for early implementation of design improvements that reduce or eliminate failure modes, improve performance, and reduce the life cycle cost of the materiel.

2. FACILITIES AND INSTRUMENTATION.

Facilities and instrumentation required to conduct reliability or durability testing need to be tailored to represent the operational and environmental conditions expected to be encountered by the item in the field. Due to the wide variety of commodity items covered by this document, it is not practical to list all possible facilities and equipment.

2.1 Facilities.

The facilities required will be specified in applicable Standard Operating Procedures (SOPs), TOPs, or in the test directive and/or detailed test plan.

2.2 Instrumentation.

The instrumentation required to apply the test conditions, monitor test item reaction to test conditions, and record the required data will be specified in applicable SOPs, TOPs, or in the test directive and/or detailed test plan.

3. REQUIRED TEST CONDITIONS.

a. Obtain and analyze the OMS/MP statement. If statements are unclear, request clarification from the originator.

b. Review the test directive and/or detailed test plan to ensure the item will be operated in field conditions which are as close to actual conditions as possible. If applicable, review TOP 2-2-506[1], Endurance Testing of Tracked and Wheeled Vehicles, to ensure the guidelines presented therein were utilized in developing the test plan. When the OMS/MP specifies several environments by percent, the test must be performed to the specified percentage of each environment.

c. Review the test directive and/or detailed test plan to ensure the length of the test will be adequate to address the criterion. For example, to meet a mean time between failures (MTBF) criterion with 90% confidence (one-sided limit) when the exponential assumption of failure rate is applicable, the test duration must be a minimum of 2.3 times the MTBF criterion and the test item must complete the test without a single failure. If a single failure occurs, then the test would have to be extended to 4.0 times the MTBF criterion in order to meet the requirement with 90% confidence. In other words, the planned test length should account for the confidence in the item. More failures would be expected for a redesigned item or a prototype and, therefore, a longer test should be planned for such an item. Raise any concerns to the test officer, materiel developer, Program Manager and/or evaluator.

d. Obtain and analyze the failure definition and scoring criteria document to ensure that it completely and precisely defines the various reliability failures, durability failures, and reliability incidents that could occur during the test. If not, request clarification from the originator or the official incident scoring community assigned to the test.

e. Review the System Support Package Components List (SSPCL), if provided, prior to and at the completion of testing to ensure a proper System Support Package (SSP) is being developed. Also, inventory the SSP against the SSPCL to ensure completeness.

f. Inventory the Basic Issue Items (BII) against the Basic Issue Items List to ensure completeness.

g. Review any supplied Safety Assessment Report (SAR) to ensure all known hazards are included and can be articulated to test maintenance and operator personnel and others as necessary.

h. Develop a data-collection plan that will obtain the data required to perform the analysis, setup an electronic database to store the data, and develop a means to analyze the data for conducting reliability and maintainability calculations. This includes utilizing software to conduct statistical calculations on the respective data.

4. TEST PROCEDURES.

The test procedures will be specified in the test directive or detailed test plan.

5. DATA REQUIRED.

Data must include all information normally recorded on the standard Test Incident Report (TIR) (DA Form 7492) as outlined by DA PAM 73-1[2]. The specific data required to evaluate each of the RAM-D and ILS elements is presented within the appropriate paragraphs of this document.

6. PRESENTATION OF DATA.

6.1 Reliability.

6.1.1 Introduction.

This section provides guidelines for calculating and presenting results from reliability and/or endurance tests. It applies to all items for which reliability criteria exists or can be developed, and for which a reliability and/or endurance test is required.

Reliability is defined as the probability that an item will perform its intended function for a specified interval under stated conditions. Because of its contribution to equipment availability and hence effectiveness, the reliability parameter is important in Army materiel.

An endurance test is a test which involves extended operation of one or more test items under cycles designed to simulate, under proving ground conditions, extended field-use. The endurance test is the principal means of producing data for reliability and maintainability during development tests, and also is a major source of information on maintainability and human factors.

6.1.2 Criteria.

Reliability criterion for an item is generally presented as either an acceptable MTBF or a probability of success of completing a defined mission without a failure and should include a confidence level to account for test parameters and real life conditions that affect, from a statistical standpoint, how confident one can be with the test data. Some factors that affect the statistical confidence in the test data are the length of testing conducted, the sample size under test, and the natural variation of the phenomenon being tested. It must be noted that the "T" in MTBF stands for whatever units the time is related to, such as, miles, hours, rounds, cycles, etc.

The following is an example of a reliability criterion: Verify with 80% confidence that the mean miles between failures (MMBF) is no less than 180 miles.

6.1.3 Data Required.

All data required to identify a reliability incident and the time, in terms of service life, when the incident occurred, should be recorded during testing. Data must include all information normally recorded on the standard TIR (DA Form 7492) as outlined by DA PAM 73-1. This includes:

(1) Descriptions of findings from all maintenance inspections and/or troubleshooting tasks.

(2) Total accumulated test item operating time in terms of hours, miles, cycles (as appropriate), courses, speed, loading factors, environment, and other operating conditions.

(3) A description of each failure. This may include:

(a) Failing component and serial number.

(b) Accumulated operating time for the test item and failing component.

(c) Seriousness of the failure; effect of the failure with regard to overall test item effectiveness and downtime.

(d) Reason for the failure, if available.

(e) Manner in which the malfunction was detected and the methods used to locate the failing component.

(f) Corrective action performed.

(g) Any recommendations for preventive maintenance.

(h) Any safety hazards associated with the failure and/or corrective action.

(i) A description of modifications and maintenance performed, including time of occurrence, maintenance times, circumstances, any special tools unique to the test item, and parts and labor expended.

6.1.4 Test Findings.

a. The test report should include the following, if applicable:

(1) Summary table(s) of the testing completed with each test item in miles, hours, cycles, etc.

(2) Tables and/or figures displaying any other important parameters that were recorded during the test, such as vehicle speed, vehicle configuration, course conditions, ride quality metrics.

(3) Summary table(s) of failures listed by Failure Definition/Scoring Criteria (FD/SC) chargeability for each test item as shown in the example table (Table 1). Chargeability assigns the primary cause for the occurrence of the test incident to one of several elements. Analysis of developmental testing reliability data is primarily focused on assessing the reliability of the system under test and does not typically emphasize operational types of failures such as those that result from crew/operator or maintenance personnel error or failures within the supply support system. As a result, common chargeabilities evaluated against the reliability criteria during developmental testing typically include, but are not limited to, hardware/CFE and software/CFE (CFE – contractor furnished equipment). Most other chargeabilities, commonly referred to as "operational" chargeabilities, such as maintenance personnel, crew/operator error, technical publications, supply support, training, and government furnished equipment, are not typically evaluated against the reliability criteria during developmental testing but are reported for informational purposes.

TABLE 1. RELIABILITY INCIDENTS BY CHARGEABILITY

Item No.	OMF[a]	EMA[b]	UMA[c]
Hardware/CFE			
A1	0	6	14
A2	1	4	14
Combined	1	10	28
B1	3	12	19
Operator/Crew			
A2	0	1	1
B1	0	0	1
Maintenance Personnel			
A1	1	1	1

[a] OMF = Operational Mission Failure.
[b] EMA = Essential Maintenance Action.
[c] UMA = Unscheduled Maintenance Action.

(4) A summary of reliability incidents per subsystem as shown in the example table (Table 2) and graphs (Figure 1 and Figure 2). Individual subsystems could be broken down by black box codes, if applicable, on a case-by-case basis if doing so would add clarity to the analysis. The table(s) and any associated figures should be prefaced by a paragraph stating which FD/SC chargeability (ies) is/are presented and, if necessary, why. The paragraph should also include an explanation of how each failure score counts toward the failure totals because some types of incidents are counted towards multiple totals. For example, if an FD/SC that includes Operational Mission Failures (OMF), Essential Maintenance Actions (EMA), and Unscheduled Maintenance Actions (UMA) is used, every OMF that occurs necessitates an essential maintenance action that is also an unscheduled maintenance action to repair the item to mission-ready status. Therefore, the following statement would be added to the report: It should be emphasized that every OMF was an EMA as well as a UMA. Likewise, every EMA was also a UMA. Similarly, if an FD/SC that includes System Aborts (SA), Essential Function Failures (EFF) and Non-Essential Function Failures (NEFF) is used, every SA that occurs was due to an EFF but cannot be due to a NEFF. Therefore, the following statement would be added to the report: It should be emphasized that every SA was an EFF.

TABLE 2. RELIABILITY INCIDENTS PER SUBSYSTEM

Subsystem	ITEM A1			ITEM A2			ALL		
	OMF	EMA	UMA	OMF	EMA	UMA	OMF	EMA	UMA
Accessories	0	0	0	0	0	1	0	0	1
Body/cab & hull	0	0	0	0	0	1	0	0	1
Brakes	0	2	5	1	1	5	1	3	10
Electrical system	0	0	1	0	0	1	0	0	2
Frame assembly	0	1	1	0	0	0	0	1	1
Springs/shocks	0	1	3	0	2	4	0	3	7
Wheels and tires	0	2	4	0	1	2	0	3	6
Total	0	6	14	1	4	14	1	10	28

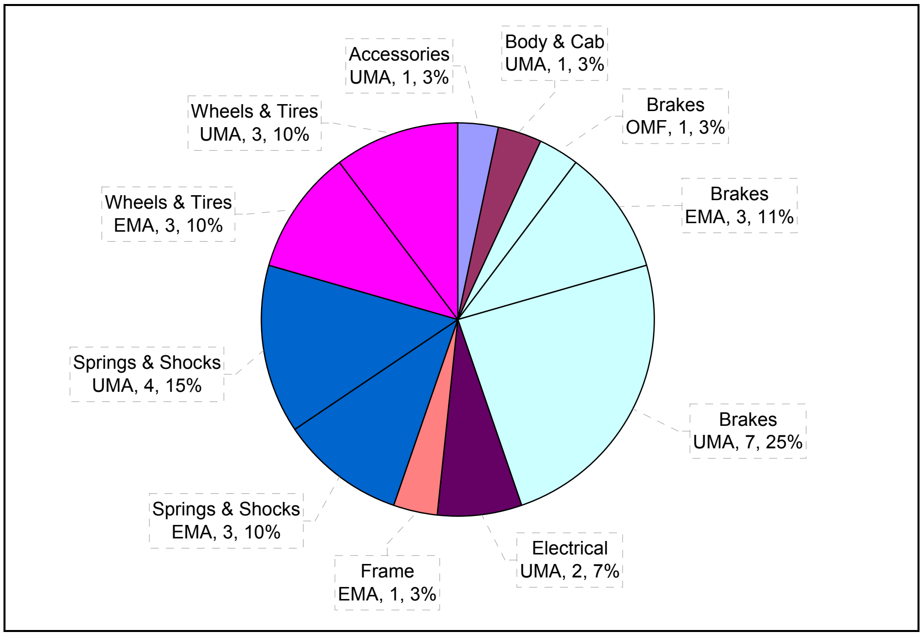

Figure 1. Percentage of reliability incidents by subsystem for items A1 and A2 combined.

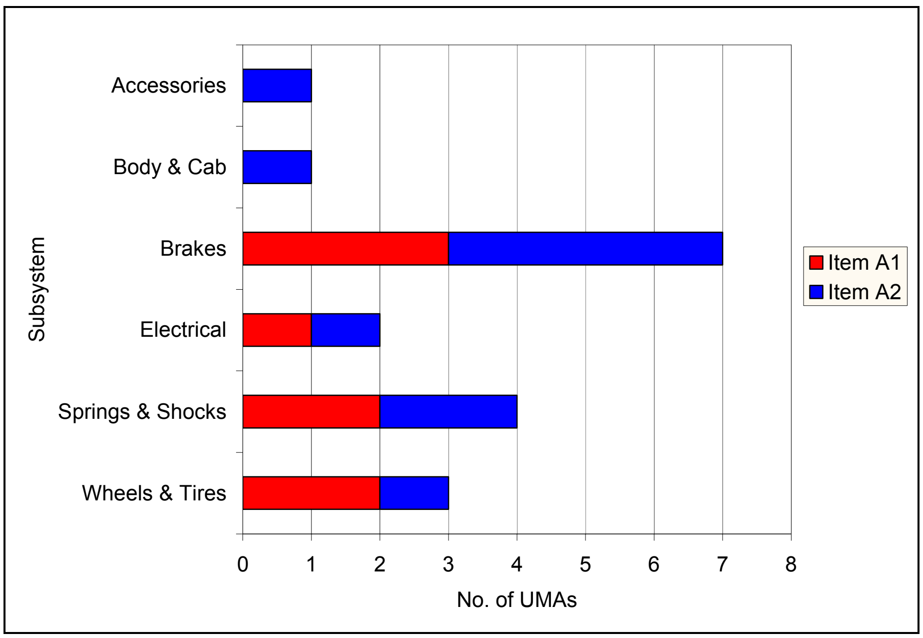

Figure 2. Combined total EMAs by subsystem for items A1 and A2.

(5) A graph showing the reliability incidents accumulated on each item compared to the reliability criteria is shown in the example plot below (Figure 3). If the test involves multiple like-model items and combination of the reliability results is statistically acceptable (see methodology in Appendix A, starting on page A-4), an additional graph could be presented showing either an average of the test results (Figure 4) or an accumulation of the test results (Figure 5). These two graphs were created using the same data as Figure 3. For tests which accumulate very large numbers of failures (hundreds/thousands), charts such as those shown in Figure 3 through Figure 5 may be difficult or impossible to compile. The figures are followed by some comments regarding additional information that can be added if desired.

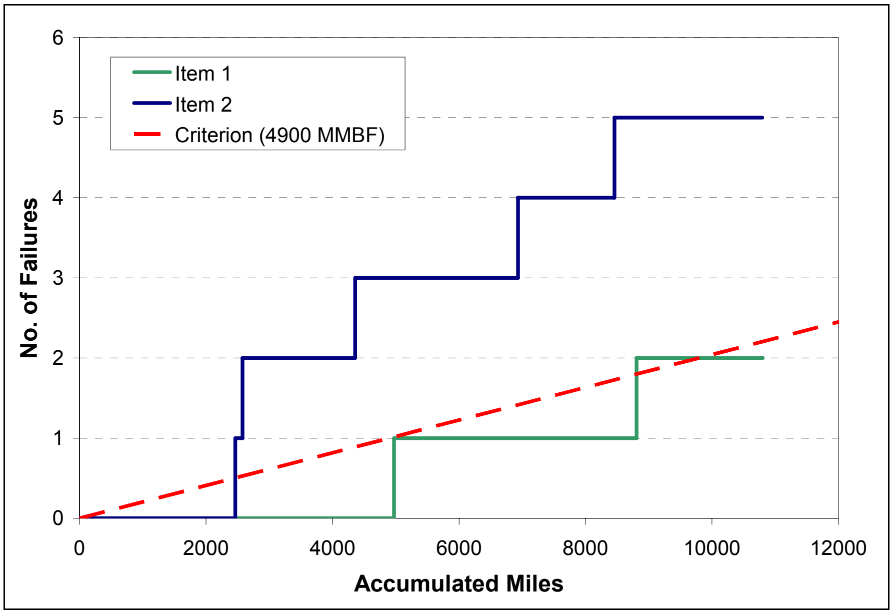

Figure 3. Failures accumulated during testing plotted against the reliability criterion.

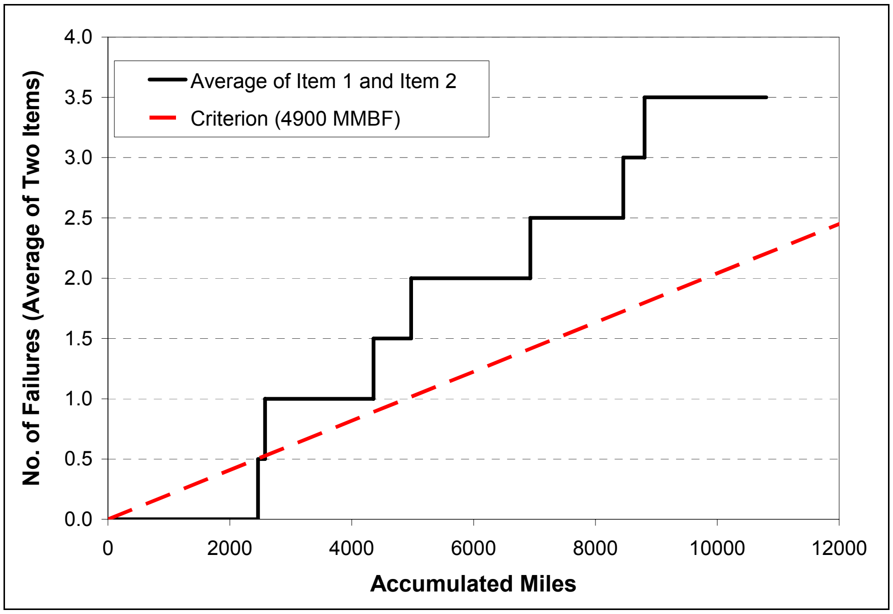

Figure 4. Averaged failure results plotted against the reliability criterion.

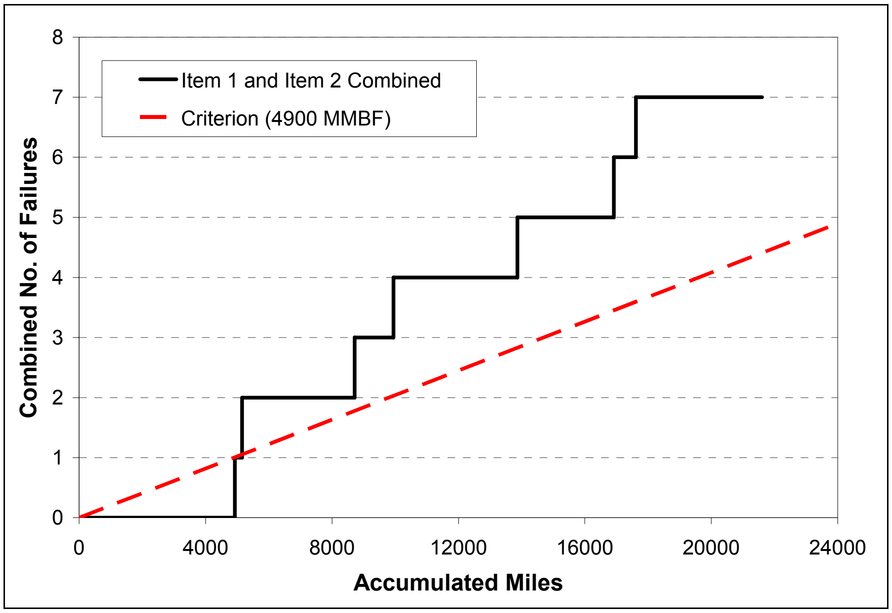

Figure 5. Combined failure results plotted against the reliability criterion.

Notes:

(a) If the test involves multiple like-model items and combination of the reliability results is statistically acceptable (see methodology in Appendix A, starting on page A-4), then the accumulated failure plot (Figure 3) could include the combined results.

(b) If testing directly follows the OMS/MP of the item being tested, additional information can be added to Figure 3 depicting the initiation/conclusion of each mission.

(c) The chart(s) may need to indicate when modifications were installed during the test to delineate between failure characteristics of the test item before and after a modification was installed.

(6) A list of the modifications performed on the test items. This table should include the modification title, the purpose of the modification, a brief description, a list of the test items that received the modification, and the test life at which the modification was applied.

(7) A summary of the major and repetitive reliability incidents that occurred during the test as shown in the example table below (Table 3). The table should include the TIR number, test item identification number, test life when the incident occurred, a brief description of the incident and the corrective action taken, the severity of the incident (incident score), and the chargeability, unless all items have the same chargeability, at which point the chargeability should be displayed in the title. The table is typically sorted by major subsystem, followed by brief description, then test item, and finally test life to show repetitive (pattern) failures. However, the evaluator should sort the data in a manner that best suits their analysis.

TABLE 3. HARDWARE/CFE RELIABILITY INCIDENTS BY SUBSYSTEM

TIR No. XX-X___	Test Item	Test Miles	Description	Score(s)-Action Performed
BRAKES				
102	Item1	5704	Service brakes out of adjustment.	EMA – Adjusted brakes.
029	Item2	1171		
031	Item2	1488		
048	Item2	2930		OMF – Replaced brake shoes.
104	Item4	6806		EMA – Adjusted brakes.
077	Item2	4477	Breakaway lever inoperative.	EMA – Replaced spring.
ELECTRICAL SYSTEM				
076	Item2	4477	Marker light inoperative.	UMA – Replaced lamp.
080	Item3	4739		
082	Item1	3973	Right taillight inoperative.	
SPRINGS/SHOCK ABSORBER				
085	Item1	4450	Left shock class I oil leak.	EMA – Replaced shock.
093	Item2	6191		
108	Item2	7500		
086	Item1	4450	Right shock class I oil leak.	
064	Item2	3921		
105	Item3	7244	Right shock class III oil leak.	

(8) Present the reliability characteristics that the test community and evaluator deemed appropriate for the system(s) under test. Table 4 is shown as an example. Statistical presentations of reliability testing provide a statement in the form of point estimates and either a lower confidence limit (LCL) or a confidence interval on reliability. One factor complicates the analysis of results from many reliability tests: usually only small sample sizes are possible because the endurance and/or reliability testing requires extensive investments of both time and money. Appendix A provides the procedures for analytical assessment of reliability. Several examples are provided to cover different situations regarding sample sizes, statistical differences, and confidence levels. Examples also provide the proper verbiage for reporting reliability results. The reliability characteristics should be prefaced by a paragraph explaining the assumptions involved and the calculations performed based on these assumptions. Here is an example: It is assumed that the time between failures is an exponentially distributed random variable.

TABLE 4. RELIABILITY CHARACTERISTICS

Test Item	Miles	Failures	MMBF Point Estimate	MMBF 80% LCL	EMA	MMBEMA Point Estimate	MMBEMA 80% LCL
HARDWARE/CFE							
Item1	7,502	0	_[a]	4,661	15	500	390
Item2	7,500	1	7,500	2,504	18	416	332
Combined	15,002	1	15,002	5,010	33	454	386
OPERATIONAL							
Item1	7,502	1	7,502	2,505	16	468	368
Item2	7,500	1	7,500	2,504	20	375	303
Combined	15,002	2	7,501	3,505	59	416	357

[a] The point estimate cannot be calculated when the number of failures is zero.

6.1.5 Technical Analysis.

a. Data shall be compared with reliability characteristics requirements as specified in the appropriate criteria. This data could be tabulated or presented in paragraph form, depending on the amount of data being analyzed. Statistical comparisons should be used to show differences between multiple like or unlike test items. For example, 1st generation versus 2nd generation results or results from different test phases (e.g. Production Verification Test (PVT) vs. Follow-on Production Test (FPT)). Also, individual failure modes and/or subsystem failures can be compared from item to item. Appendix A provides the procedures for statistical comparison of reliability results. The results of a statistical comparison should be reported in the following manner: A statistical test was performed to determine if the reliability characteristics of same-model test items could be combined. No significant differences were detected at the 10% significance level (or other acceptable level); therefore, the results were combined and the cumulative failure rate was compared to the reliability criterion.

b. The report should include an analysis of all significant and/or repetitive failure modes.

c. The presentation shall conclude with a summarization of the adequacy of the test item reliability characteristics and any recommendations for improvement.

6.2 Durability.

6.2.1 Introduction.

a. Durability is a special case of reliability. Durability is defined as the probability that an item will successfully survive its projected life, overhaul point, or rebuild point without a durability failure. A durability failure is considered to be a malfunction that precludes further operation of the item and is great enough in cost, safety, or time to restore, that the item or major subsystem (e.g. engine or transmission) must be replaced or rebuilt.

b. A durability test relates to the mathematical probability that a vehicle or a major component thereof will be able to operate under defined conditions for a specified time or distance before requiring major overhaul, replacement, or salvage. A complete durability evaluation is seldom performed on developmental vehicles since large sample sizes, extensive test mileage, and considerable funding are required. When a durability test is conducted, reliability characteristics of the item are also evaluated.

6.2.2 Criteria.

A durability criterion for an item is generally presented as a probability of successful completion of a specified number of kilometers, miles, or hours without requiring major overhaul, replacement, or salvage (e.g. Each item shall have a 0.6 probability with a 50% confidence of completing 20,000 mi. (32180 km) per the mission profile without a durability failure).

6.2.3 Data Required.

Same as reliability (see Paragraph 6.1.3).

6.2.4 Test Findings.

a. Same as reliability (see Paragraph 6.1.4).

b. Statistical presentations of durability testing provide a statement in the form of point estimates and a LCL on reliability. Two factors complicate the analysis of durability test results: first, usually only small sample sizes are possible because the durability testing requires extended periods of time and destroys the tested items; second, durability failures typically occur at the end of the service life of the item and very little is usually known or can be assumed about the statistical distribution of durability failures. Appendix B provides the procedures for analytical assessment of durability. Thirteen examples are provided to cover different situations regarding sample sizes, mission probabilities, and confidence levels. Examples also provide the proper verbiage for reporting durability results.

6.2.5 Technical Analysis.

a. Data shall be compared with durability characteristics requirements as specified in the appropriate criteria.

b. The report should include an analysis of all significant and/or repetitive failure modes.

c. The report shall conclude with a summarization of the adequacy of the test item durability characteristics and any recommendations for improvement.

6.3 Integrated Logistic Supportability.

6.3.1 Introduction.

a. Integrated logistics support, often called system support and sustainment, refers to the materiel and services required to enable the operating forces to operate, maintain, and repair the end item within the maintenance concept defined for that end item. Logistics support encompasses the identification, selection, procurement, scheduling, stocking, and distribution of spares, repair parts, facilities, support equipment, trainers, technical publications, contractor engineering and technical services, and personnel training as necessary to keep the end item in a functioning status. Integrated logistics supportability is an attribute of a particular materiel system that encompasses these factors.

b. The test directive and/or detailed test plan indicates the scope of the logistic supportability subtest to be conducted. The following sub-elements are provided as an outline of the considerations that would, in general, cover the subject of logistic supportability. This outline must be tailored to fit the test effort being conducted and the system under test. Additional sub-elements may be included depending upon the complexity of the test effort and the system under test. Criteria for the sub-elements must be extracted from program documentation, such as requirements documents, acquisition plans, and materiel specifications. The typical logistic supportability sub-elements are as follows:

(1) Maintainability Indices.

(2) Supply Support.

(3) Technical Data/Equipment Publications.

(4) Support and Test Equipment.

(5) Training and Training Devices.

(6) Design for Maintainability and Safety and Human Factors during Maintenance.

Other sub-elements of logistic supportability that are rarely addressed during testing include: Facilities; Packaging, Handling, Storage, and Transportation; Manpower and Personnel.

6.3.2 Maintainability Indices.

a. Introduction.

Maintainability is defined as a characteristic of design and installation which provides inherently for the item to be retained in or restored to a specified condition within a given time, when the maintenance is performed in accordance with prescribed procedures and resources.

Maintainability indices are the measures by which the maintainability of a system/item is quantitatively assessed. The following is a list of the indices that should be included, if applicable, in an ILS report:

(1) Achieved Availability.

(2) Inherent Availability (A_i).

(3) Maintenance Ratio (MR).

(4) Mean Time to Repair (MTTR).

(5) Maximum Time to Repair (MaxTTR) (at a specified percentile).

(6) Probability of Organizational (or other required level) Maintenance.

b. Criteria.

Typical maintainability criteria statements are as follows:

(1) Each model shall have a MR no greater than 0.0050 maintenance man hours per operating mile.

(2) The mean time to repair at the unit level of maintenance shall not exceed 1.0 hours. The 90th percentile maximum time to repair at the unit and direct support maintenance levels shall not exceed 1.25 and 2.25 hours, respectively.

c. Data Required.

All maintenance operations required during the course of testing will be performed in accordance with the applicable technical manuals (TMs). When applicable, the TIR narrative should include notification that manufacturer-supplied, sponsor-approved information is supplementing/updating/correcting information contained in the TMs. Preventive maintenance will be performed at the specified intervals. All maintenance actions will be monitored to accumulate required data. The following maintenance data will be acquired and recorded in TIRs:

(1) Each scheduled and unscheduled maintenance action, maintenance task(s) performed, total man-hours and clock hours expended for each maintenance level, maintenance chargeability, and subsystem that required maintenance.

(2) Vehicle component failures, including maintenance-related components and systems, will be monitored throughout testing. Any malfunctions or problems encountered while performing maintenance tasks will be recorded. At the time of each malfunction, the failed component or assembly will be identified, and the accumulated operating time (hours and miles) of the system and life (hours and miles) of the failed component will be determined and recorded.

d. Test Findings.

(1) Hardware and operational maintenance data summaries. An example summary is located in Table 5. These tables should include total maintenance tasks, clock hours and man hours accumulated on each test item broken down by type (scheduled, unscheduled, on-condition, etc.) and level (unit, direct support, general support, etc.). The maintenance summary tables should also include all applicable maintainability indices. The indices are computed as follows:

(a) Achieved Availability:

$$A_a = \frac{\text{Operating Hours}}{\text{Operating Hours} + \text{Total Maintenance Clock Hours (Scheduled and Unscheduled)}}$$

Or, equally:

$$A_a = \frac{\text{MTBM}}{\text{MTBM} + \overline{\text{M}}}$$

Where,

MTBM = mean time between maintenance actions
$\overline{\text{M}}$ = mean maintenance downtime.

(b) Inherent Availability:

$$A_i = \frac{\text{Operating Hours}}{\text{Operating Hours} + \text{Total Unscheduled Maintenance Clock Hours}}$$

Or, equally:

$$A_i = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

(c) Maintenance Ratio:

$$\text{MR} = \frac{\text{Total Maintenance Man - Hours}}{\text{Total Operating Hours}}$$

Maintenance ratio criteria are often specified in terms of maintenance man-hours per operating mile. When this is the case, be sure to report the units so the reader does not assume the maintenance ratio is dimensionless (same units in both numerator and denominator).

(d) Mean Time to Repair:

$$MTTR = \frac{\text{Total Unscheduled Maintenance Clock Hours}}{\text{Total Unscheduled Maintenance Tasks}}$$

(e) Maximum Time To Repair at the 90th Percentile (MaxTTR 90%) is the corrective maintenance clock hours corresponding to the 90th Percentile task number (Typically, 90th percentile is calculated, but 95th percentile is also commonly required).

(f) Probability of Org Maintenance:

$$\text{Probability of Org} = \frac{\text{Total Crew and Organizational Maintenance Tasks}}{\text{Total Maintenance Tasks for All Maintenance Levels}}$$

(2) A summary of active hardware-chargeable maintenance hours and tasks expended per subsystem on each test item and/or combined. An example table and graph are located in Table 6 and Figure 6, respectively.

TABLE 5. MAINTENANCE SUMMARY

Characteristic	Item1	Item2	Item3	Item4	Combined
Total miles	7,502	7,500	7,501	7,501	30,004
Total operating hours	375.1	375.0	375.1	375.1	1,500.2
Total maintenance tasks	15	16	10	9	50
Scheduled, org	1	1	1	1	4
On-condition, scheduled, org	2	3	1	1	7
Unscheduled, org	12	12	8	7	39
Total active clock hours	5.6	12.6	7.7	9.9	35.8
Scheduled, org	1.5	3.0	3.4	3.7	11.6
On-condition, scheduled, org	1.9	3.8	1.7	1.8	9.2
Unscheduled, org	2.2	5.8	2.6	4.4	15.0
Total active man-hours	5.6	14.2	8.2	10.2	38.2
Scheduled, org	1.5	3.3	3.4	3.7	11.9
On-condition, scheduled, org	1.9	4.6	1.8	1.8	10.1
Unscheduled, org	2.2	6.3	3.0	4.7	16.2
Maintenance Ratio, org	0.0008	0.0019	0.0011	0.0014	0.0013
Mean Time To Repair, hr	0.19	0.48	0.33	0.63	0.39
Max Time To Repair (90%), hr	0.45	0.88	0.57	1.25	0.66
Probability of Org Maintenance	1.00	1.00	1.00	1.00	1.00
Achieved Availability (A_a)	0.98	0.97	0.98	0.97	0.98
Inherent Availability (A_i)	0.99	0.98	0.99	0.99	0.99

TABLE 6. HARDWARE CORRECTIVE MAINTENANCE PER SUBSYSTEM

Subsystem	Item1		Item2		Item3		Item4		ALL	
	Man-Hours	Tasks	Man-Hours	Tasks	Man-Hours	Tasks	Man-Hours	Tasks	Man-Hours	Tasks
Accessory Items	0	0	0	0	0.02	1	0	0	0.02	1
Body & Cab	0	0	0.2	1	0	0	0	0	0.2	1
Brakes	0.4	3	3.9	4	1.8	2	3.4	3	9.5	12
Electrical System	0.1	1	0.1	1	0.1	1	0	0	0.3	3
Frame Assembly	0.2	1	0	0	0	0	0.6	1	0.8	2
Springs & Shocks	1.2	3	0.8	4	1.1	4	0.2	1	3.3	12
Wheels & Tires	0.3	4	1.3	2	0	0	0.6	2	2.2	8
Total	2.2	12	6.3	12	3.0	8	4.8	7	16.3	39

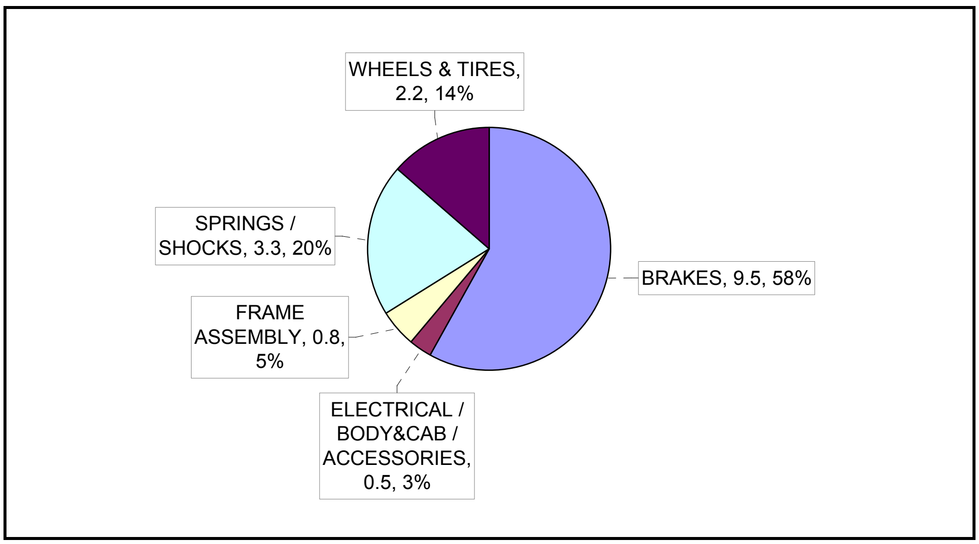

Figure 6. Percentage of unscheduled hardware maintenance man-hours by subsystem for all test items combined.

(3) If there is a MR requirement, provide a plot of the accumulated maintenance man-hours per mile versus the MR requirement as shown in the example plot below (Figure 7).

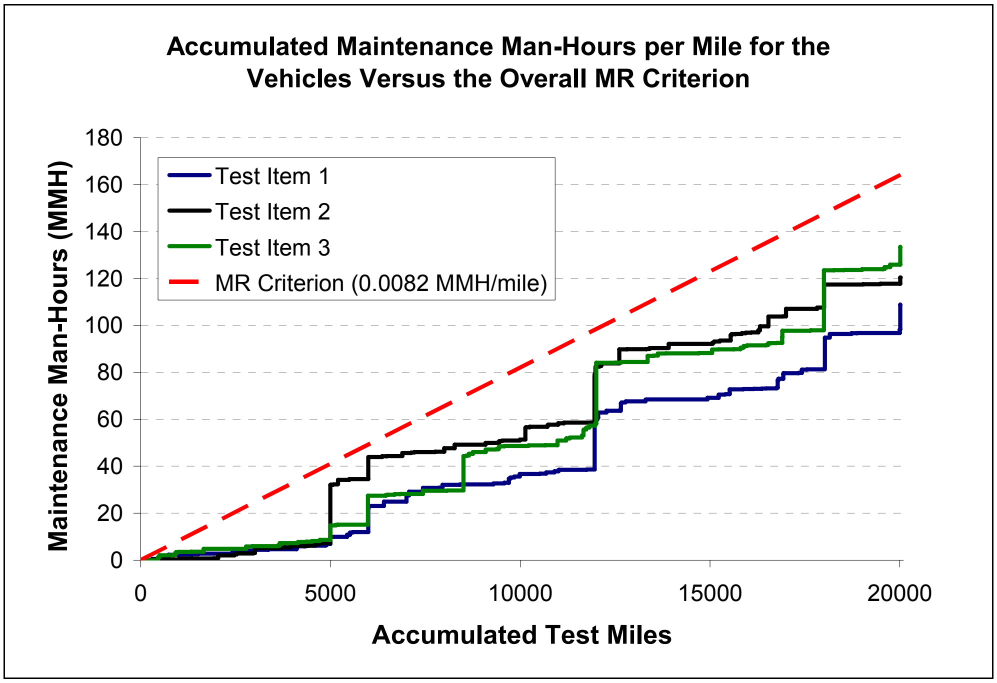

Figure 7. Accumulated maintenance man-hours per mile plotted against the MR criterion.

(4) An appendix containing the Supportability Analysis Chart (SAC), if requested. The SAC must include the following columns of data:

(a) Group and sequence numbers. Enter the functional group number as indicated in the Maintenance Allocation Chart (MAC) of the assembly or subassembly. Enter the sequence number of the maintenance task in parentheses below the group number. The task which caused the end item to undergo maintenance should be the first in the sequence. Use a number to denote a layover. Use a letter to denote the separate tasks if more than one is performed during the same layover; e.g., 1a, 1b, 1c, 2, 3, 4a, 4b, 5, etc.

(b) Component and related operations. Enter component and related maintenance functions as indicated in the MAC. Maintenance functions assigned to depot category maintenance are not normally shown.

(c) Maintenance category, prescribed. Indicate the maintenance category prescribed using the single-character code from the MAC (e.g. C – operator/crew; O – organizational; F – direct support; H – general support; N – none prescribed).

(d) Maintenance category, recommended. Indicate the maintenance category recommended by the installation/field operating activity using a single-character code as above.

(e) TM instructions, adequate. Place an “X” in this column to indicate that the TM instructions covering this maintenance task or action are adequate.

(f) TM instructions, inadequate. When the TM instructions are considered inadequate, insert the installation/field operating activity TIR number, if appropriate, which transmitted the DA Form 2028.

(g) Active maintenance time. Enter clock hours and man-hours required for the maintenance operation to the nearest tenth of an hour. If the operation was not actually performed, indicate the estimated active maintenance time by using the prefix E. (Explain unusual differences in maintenance times for the same operation in the body of the TIR.)

(h) System life. Enter the number of operational hours, miles, rounds, cycles, events, etc., as required in the test plan, accumulated during the test before the malfunction or scheduled service occurred. Under the life figure, enter in parentheses the sequence number for which that particular operation was last performed on that particular end item. Place an “S” in this column if the operation was performed on a sampling basis and not because of an actual maintenance action.

(i) Reason performed. Enter the symbol “Unsched” if this operation was performed as a result of unscheduled maintenance. If the operation was performed and recorded as a required portion of a scheduled maintenance service, enter the symbol “Sched.” If the operation was performed only to verify procedures or tool requirements, not to correct a malfunction, enter the symbol “Sim” or “Demo.”

Note: Use a separate SAC to record each simulated maintenance action.

(j) Remarks. When a maintenance task is related to a specific incident, enter the TIR number. Also, use the remarks column to identify maintenance functions which are considered failures for reliability computations. Enter the time in man-hours prescribed by the MAC to perform each function (or locally devised forms may require entry of the information in a separate column).

Note: For the test report, include the serial number of the test item in the remarks column or include separate charts for each test item.

e. Technical Analysis.

(1) Data shall be compared with maintainability indices requirements as specified in the appropriate criteria.

(2) The report shall conclude with a summarization of the adequacy of the test item maintainability characteristics and any recommendations for improvement.

6.3.3 <u>Supply Support.</u>

a. Introduction.

The support system data are vital in determining the maintainability of the overall item/system. The substitution of unauthorized repair parts or the procurement of repair parts other than through normal channels supporting the test item is not authorized. Parts approved by the Program Manager (PM) and agreed upon by the evaluator may be acquired locally by the manufacturer field service representative (FSR) or the Test Center "as necessary" or "to expedite testing." The maintenance historical data generated during the test, particularly parts consumption data, will materially assist provisioning personnel in determining the required logistic support for the item/system.

b. Criteria.

A typical supply support criteria statement is as follows:

Repair parts shall be authorized in adequate quantities and diversity at the appropriate levels, consistent with the MAC, Repair Parts Special Tools List (RPSTL), and skills required to install and align the parts. Repair parts which are used to maintain the system must be interchangeable with like parts being replaced.

c. Data Required.

(1) Throughout testing, all maintenance operations requiring parts replacement shall be observed and the following data recorded in TIRs:

(a) Nomenclature, functional group number, and national stock number (NSN) or manufacturer's part number of all parts used, when provided.

(b) Maintenance category prescribed by the RPSTL and/or the MAC, when provided.

(c) The part life (miles, operating hours, and/or operating cycles) of the replaced item.

(d) The reason for replacing the part (scheduled or unscheduled).

(e) Comments on any difficulties in installation, alignment, or interchangeability of parts.

(f) A record of repair parts that were not compatible with the test item or not prescribed at the correct maintenance level, when provided.

(2) Repair parts shall be examined with respect to the maintenance category authorized to stock and/or requisition the part and the category prescribed to replace the part, when provided.

(3) Standard parts should be used when possible. Peculiar parts will be examined to determine if they can be replaced with standard items already within the logistics system.

(4) Repair parts will be examined to ensure modular design has been considered.

(5) Repair parts used on the test item will be compared with the repair parts manual to ensure that the data in the manual are adequate to permit identification/acquisition of the parts by personnel in the field.

d. Test Findings.

(1) List and describe any recorded discrepancies with the repair parts used during the test. For example, list all unserviceable repair parts received during testing and explain each discrepancy and corrective action performed.

(2) Identify the total number of repair parts used.

(3) Provide a table listing all parts replaced during the test. This table should include the part number, nomenclature, and the total number of each part replaced.

(4) Provide an appendix containing the Supply Support Chart (SSC), if requested. The SSC must include the following columns of data:

(a) Group and sequence numbers. Indicate parts usage by maintenance operation by a cross-reference to the group number and sequence number from column 1 of the SAC.

(b) NSN. Enter one of the following: NSN, manufacturer's part number, or drawing number (in this order of preference).

(c) Noun nomenclature. Enter the nomenclature as listed in the parts manual.

(d) Maintenance category, prescribed. Indicate the maintenance category prescribed by the parts list under review using the Source, Maintenance, and Recoverability (SMR) codes from the TM. (Use code N if none prescribed)

(e) Maintenance category, recommended. Use the SMR codes to indicate the maintenance category recommended by the installation/field operating activity.

(f) Part life. Enter the number of operating hours (essential), miles, rounds, events, etc., as required by the test plan accumulated by this part. This is the actual part life and should agree with the part life reported on the TIR. After each entry in this column, enter the appropriate life unit symbol; i.e. H, M, or R.

(g) Reason used. Enter the symbol "Unsched" if this part was used as a result of unscheduled maintenance. If the part was replaced as a required action of scheduled maintenance, enter the symbol "Sched." If the part was consumed to verify procedures or tools, not to correct a malfunction, enter the symbol "Sim."

(h) Remarks. If the part is related to a specific incident, insert the TIR reference number. This column may also be used to identify parts used to correct failures or to list a quarterly service.

e. Technical Analysis.

The test report should include comments as to the qualitative effects of supply support inadequacies on the maintenance calculations. For example, installation of incorrect or nonfunctioning repair parts affects availability.

6.3.4 Technical Data/Equipment Publications.

a. Introduction.

The technical data/equipment publications sub-element is conducted to ensure the publications Interactive Electronic Technical Manuals (IETM), and/or TM-equipped PDAs completely and adequately reflect the system they support. Technical data/equipment publications must be easily and completely understood by the maintenance personnel to whom they are addressed. Each manual, operator level through general support level, must be compared with the test item and with the specifications and military standards prescribing format, technical content, and standard of production for TMs (MIL-STD-38784[5] and MIL-M-63000™ series[6]). Comments, as appropriate, will be made to the preparing agency with information copies to Army Materiel Command (AMC) and the appropriate US Army Training and Doctrine Command (TRADOC) agency.

b. Criteria.

A typical technical data/equipment publications criteria statement is as follows:
All technical documentation, including MAC, training materials and operator/maintenance manuals must be provided to field a total system. Equipment publications contained in the SSP must be complete, accurate, easy to read, simple to follow, and adequate to permit accurate operation and completion of both scheduled and unscheduled maintenance operation and identification of parts at all levels of maintenance. Also the lubrication intervals shall be as required by the component manufacturers. A lubrication chart shall be furnished with each vehicle and shall include interchangeable military lubricants as well as commercial lubricant designations.

c. Data Required.

Any problems noted with the technical data/equipment publications will be recorded in TIRs. Manuals will be reviewed to determine the following:

(1) Simplicity, clarity, and completeness of the manuals to include whether operating instructions are commensurate with the training and skill of the operator/crew.

(2) Adequacy of troubleshooting instructions, as applicable; completeness of preventive maintenance procedures; and adequacy of safety instructions for personnel and equipment, including environmental protection during operation and maintenance.

(3) Adequacy of lubrication charts and orders with emphasis on clarity and completeness. Determine if specified lubricants are in the Army supply system.

(4) Errors and omissions in nomenclature and stock numbers on repair parts and special tool kits.

(5) Adequacy of instructions for the level of training and skill possessed by maintenance personnel. Record the need for additional or special training requirements.

d. Test Findings.

(1) List and describe any recorded discrepancies with the technical data/equipment publications used during the test.

(2) Provide an appendix containing the Technical Data/Equipment Publications (TDEP) Chart, if requested. The TDEP must include the following columns of data:

(a) Number. Enter the Army or manufacturer's publication or draft manual number.

(b) Quantity. Enter the number of copies received. Insert "0" if none were supplied. Publications contained in the SSP should cover operations and functions through general support maintenance and should specify the categories involved.

(c) Title. Enter the complete title.

(d) Date received, literature. Enter the date the publication was received.

(e) Date received, materiel. Enter the date the test item or materiel was received.

(f) Evaluation, adequate. Insert "X" if the publication was deemed adequate.

(g) Evaluation, inadequate. Insert "X" if the publication was deemed inadequate. Minor errors noted on DA Form 2028 are not sufficient reasons to term a publication inadequate.

(h) DA Form 2028. Insert Environmental Program Requirements (EPR) number, if appropriate, and the date the DA Form 2028 was forwarded.

(i) Remarks. In addition to remarks describing errors or omissions, explain if the publication was not evaluated and the reason therefore.

e. Technical Analysis.

The test report should include comments as to the qualitative effects of inadequacies in technical data/equipment publications on the maintenance calculations. For example, troubleshooting inadequacies would affect availability in that contractor logistic support (CLS) would be required.

6.3.5 Support and Test Equipment.

a. Introduction.

This sub-element is conducted to determine the adequacy of support and test equipment planned for use in support of the test item in a field environment. The sub-element will address the adequacy of support and test equipment listed in the SSP, plus common items not included in the SSP but available at the test site which are planned for use in support of the item when fielded. Maintenance is performed on the test item using the support and test equipment provided in the SSP following the instructions provided in the maintenance instructions.

b. Criteria.

A typical support and test equipment criteria statement is as follows:

Built-in Test Equipment (BITE) is required for common modes of failure which shall be identified on a within-the-cab display providing for rapid operator/maintainer actions. Data bus communications for electronic controlled drive train components shall be in accordance with SAE J1939[7], diagnostics shall be in accordance with SAE J1708[8] and easily upgradeable by software. Information diagnostics as available from drive-train Electronic Control Units (ECUs) and Electronic Control Modules (ECMs) or SAE J1708 data bus shall be accessible at the on board diagnostic connector. The connector for the diagnostic shall be a J1939 Deutsch 9 pin female connector that handles both J1708 and J1939 protocols. The connector shall include a cap that keeps the connector dry from water and moisture.

c. Data Required.

During the test, record the following in TIRs for system-peculiar support and test equipment.

(1) Nomenclature, NSN or part number, and date received of each special tool and each item of test equipment.

(2) TM in which the special tool or test equipment was listed.

(3) Maintenance category prescribed and recommended for use of each tool and item of test equipment when provided.

(4) Adequacy of special tools and test equipment in performance of their intended tasks.

(5) Adequacy of TM instructions for use of tools and test equipment when provided.

(6) If special tools and/or test equipment could have been replaced with common tools.

(7) Requirement for additional tools or test equipment not identified prior to testing.

d. Test Findings.

(1) Support and test equipment which is required but not available in the SSP will be identified and reported in the test report.

(2) The test report should contain discussions of any problems associated with the use of common support and test equipment with respect to the test system.

(3) Provide an appendix containing the Support and Test Equipment Chart (STEC), if requested. The STEC must include the following columns of data:

(a) Nomenclature or description. Enter the nomenclature as shown in the manual or, if none, enter the noun nomenclature and a brief description of the item. (Enter in parentheses the number of like items received, such as "(2ea)".)

(b) NSN or part number. Enter one of the following: NSN, part number, or drawing number (in this order of preference).

(c) Maintenance category, prescribed. Indicate the maintenance category authorized for the item as prescribed by the technical publication.

(d) Maintenance category, recommended. Indicate the maintenance category that the installation/field operating activity recommends for the item. If the item is not required, enter "None."

(e) Date received. Enter the date the item was received (e.g. "6/08"). Enter "Not rec" if the item was not received.

(f) Evaluation, adequate. Enter an "X" if the STEC was deemed adequate. Make no comment on items marked "None" in column d.

(g) Evaluation, inadequate. Enter an "X" if the STEC was deemed inadequate. Make no comment on items marked "None" in column d.

(h) Required, yes or no. Enter a "Yes" to indicate that the equipment is required at the maintenance category indicated in column d. Enter a "No" to indicate that the equipment is not required. This column should be marked "No" when "None" is marked in column d.

(i) TM in which listed. Enter the number of the technical publication for the test item in which the equipment is listed.

(j) Remarks. If a TIR is related to the item, enter the TIR number. Indicate if the item was used only to verify the need for the item. When an item is not required, indicate the standard item which will perform the required maintenance function, if appropriate.

e. Technical Analysis.

The test report should include comments as to the qualitative effects of inadequacies in support and test equipment on the maintenance calculations. For example, incompatible software and/or glitches render the interactive troubleshooting unusable and affect maintainability/availability indices.

6.3.6 <u>Training and Training Devices.</u>

a. Introduction.

This sub-element of the logistic supportability subtest is conducted to assess the adequacy of training provided toward performing logistic supportability activities and the ability of the training devices provided to accurately perform their function in support of maintenance training.

b. Criteria.

Typical training and training devices criteria statements are as follows:

(1) Contractor training for test personnel shall be adequate for the intended operation and maintenance support.

(2) The provided training shall be adequate to perform all operator functions and to properly maintain the end item at the unit and intermediate maintenance levels. If training devices are required, the quantities shall be adequate to support the training.

c. Data Required.

During training exercises and during testing, the following should be recorded in TIRs:

(1) Comments regarding the methods used to train, update, and familiarize the test personnel with the end item system. Comments regarding manuals and publications used and the extent of on-the-job training accomplished. Comments as to whether this instruction provided to test personnel is similar to planned field instruction in terms of content and method.

(2) Comments as to the suitability of the draft TM's for training purposes.

(3) Comments on the effectiveness of the training to perform the required operator and maintenance functions.

(4) Record of any incidents that were the result of inadequate or improper training.

(5) Record of provided training devices, if required.

(6) Comments on the provided training TMs and technical documentation.

(7) Comments on the skill level of the trainer.

d. Test Findings.

List and describe any recorded discrepancies with the training and training devices used during the test. For example, list all incidents during testing where inadequate or incorrect training caused additional troubleshooting to be performed.

e. Technical Analysis.

(1) The test reports should include comments as to the qualitative effects of inadequacies in the training and training devices on the maintenance calculations. For example, any undue difficulty encountered by the operator or maintenance personnel related to deficiencies in training would adversely affect maintainability/availability.

(2) The criteria will be considered met if the training and training devices were adequate to perform all operator functions and to properly maintain the end item.

6.3.7 Design for Maintainability and Safety and Human Factors during Maintenance.

a. Introduction.

This sub-element of the logistic supportability subtest is conducted to assess if the test items adhered to good maintainability-design principles and if safety and human factors aspects of maintenance operations were satisfactory.

b. Criteria.

Examples of typical criteria statements are:

(1) Engine Accessibility. The vehicle shall have a 0.90 probability of removing and reinstalling a replacement engine, transmission, or engine and transmission assembly, in less than 12 clock hours by no more than 2 individuals plus one crane/wrecker operator for all body styles utilizing existing Army maintenance equipment. There shall be easy accessibility to the engine for normal checks and services.

(2) Safety. Exposed components and systems which are subject to high temperatures or high pressures, are electrically actuated, or are inherently hazardous shall be provided with correct safeguarding and insulating features.

(3) The design shall provide work environments which foster effective procedures, work patterns, and personnel safety and health, and which minimize discomfort, distraction and any other factors which degrade human performance or increase error. Design shall also be directed toward minimizing personnel and training requirements within the limits of time, cost, and performance trade-off.

c. Data Required.

During conduct of the test, the following should be recorded in TIRs:

(1) Factors which establish that ease of maintenance either had or had not been included in the design.

(2) Reliability of components and other factors which indicate that the equipment design either had or had not been directed toward minimizing maintenance.

(3) Throughout all maintenance operations, observations involving the man-item relationship and safety of maintenance were recorded.

Note: AMC Pamphlet 706-134[9], Maintainability Guide for Design, can be used as a guide in determining maintainability-design characteristics. Information contained in MIL-STD-1472F[10], Human Engineering Design Criteria for Military Systems, and the Human Engineering Data Guide for Evaluation (HEDGE) can also be used for guidance.

d. Test Findings.

(1) List any findings relevant to the required criteria.

(2) List all observations related to design for maintainability, human factors and safety aspects of maintenance operations.

e. Technical Analysis.

Data shall be compared with design for maintainability and safety and human factors aspects of maintenance requirements as specified in the appropriate criteria.

APPENDIX A. RELIABILITY ANALYSIS USING THE EXPONENTIAL DISTRIBUTION

a. ANALYSIS DISCUSSION.

Reliability testing is generally carried out by operating a group of test items under conditions that duplicate field conditions as closely as possible until some predetermined test time occurs. The predetermined required test time should be designed to be some percent of the anticipated service life of the item, agreed upon by the materiel developer, Program Manager, evaluator, and the test agency.

MTBF point estimates and LCL are the typical parameters that are calculated. The LCL (80% minimum is typical) is determined typically by the application of the Chi-Squared (χ^2) distribution (See Equation (5)). If desired, the mission reliability and LCL can be calculated.

The criteria to be assessed should clearly state whether the point estimate or the LCL will be evaluated. If the criteria are unclear, then request clarification from the Program Manager or the evaluator. The point estimate is the arithmetic mean of the time between failures and equates to a very low confidence. The LCL is the best estimate of the true unknown success rate. The upper confidence limit is not generally calculated since the evaluator is typically not interested in how “good can the product be,” but rather in what is the “worst case” that can be expected.

Methodology.

The steps required to address a reliability criterion are:

(1) Assemble all failure or reliability incident data (incident scores, chargeabilities, times, etc.)

(2) Determine, if unknown, or verify the statistical distribution of the data set being analyzed. If a distribution cannot be determined, an alternative analysis method, such as a non-parametric method, can be used.

(3) Calculate the reliability parameters.

(4) Determine if the results for multiple test items can be combined.

(5) Look for statistical differences between similar items, other tests, subsystems, etc.

(6) Assess the criterion.

The following contains detailed explanations of Steps 2 through 5, followed by six examples which cover some typical reliability scenarios that can be used in assessing reliability criteria. Step 2 in a reliability analysis is verifying the validity of the assumed statistical distribution of failure times or determining the distribution, if unknown. In some cases, a graphical procedure can be used for a quick indication of the validity of the assumption. When the distribution is unknown or if the results from the graphical procedure are inconclusive for a known distribution, statistical "goodness-of-fit" tests (e.g. chi-squared, Kolmogorov-Smirnov, Cramer-Smirnov-Von-Mises, or Anderson-Darling) may be used to determine whether the assumption of a particular distribution of failure times is reasonable. If a distribution cannot be determined, an alternative analysis method, such as a non-parametric method, can be used.

Step 3 in a reliability analysis is calculating the reliability parameters. The reliability parameters should always be presented as a point estimate coupled with either a confidence limit or a confidence interval to qualify the results. The confidence level gives a statistical representation of the results that could be expected if the test were repeated. The accuracy of the confidence level, or interval, depends on the quality and quantity of the available data. One of the most common statistical distributions used for calculating reliability parameters, and the distribution covered in the following sections, is the exponential distribution.

<u>Exponential Distribution:</u>

Applicable Systems and Items: The exponential distribution is used to assess the reliability of complex systems, when the failures are due to random causes, and when the failure rate is approximately constant. Some typical systems that have been observed to meet the characteristics of the exponential distribution during their useable lives are: Wheeled vehicles, tracked vehicles, weapons, and generators.

(1) The general reliability function or probability that an item will not fail during operation until time t is:

$$R(t) = e^{-\int_0^t \lambda(t)dt} \qquad \text{(A-1)}$$

where $\lambda(t)$ is the instantaneous failure rate at time t.

For most complex systems, it has been observed that the failure rate characteristic for the equipment lifetime can be represented by the life cycle (bathtub) curve of Figure A-1.

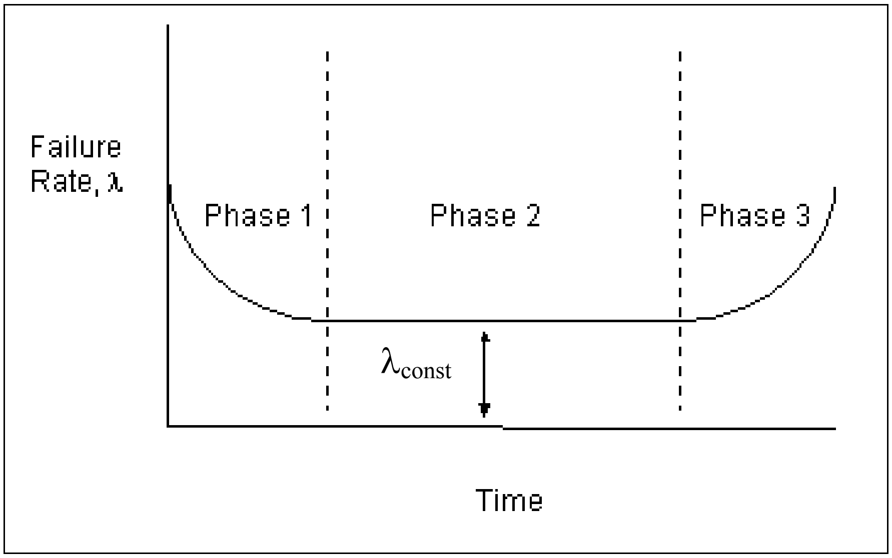

Figure A-1: Time Distribution of Complex Systems Failure Rates (Bathtub Curve)

Phase one of the life cycle curve is the early failure (infant mortality) period when high failure rate elements and workmanship defects are found. The failure rate is generally expected to decrease rapidly with time during this period.

Phase two of the life cycle curve, is the constant failure rate period during which random failures occur and is the operational life span for which reliability is generally considered. This phase is often referred to as the "usable" life because it is the period in the life of the item when the failure rate is the lowest. Since the failure rate, $\lambda(t)$, is constant during this period, the reliability function or probability (of success or survival) that no equipment failures occur during the time interval *0* to t is

$$R(t) = e^{-\lambda t} \tag{A-2}$$

The exponential characteristic given by Equation 2 is also called the exponential law.

By differentiating the cumulative distribution function (1 - R(t)), which is the probability of failure until time t, the probability density function, Equation A-3, of the exponential distribution, is obtained.

$$f(t) = \lambda e^{-\lambda t} \tag{A-3}$$

The expected time to failure or MTBF is then given by:

$$MTBF = \int_0^\infty t\lambda e^{-\lambda t}\,dt = \frac{1}{\lambda} \tag{A-4}$$

Phase three of the life cycle curve is the wear out period when some elements of the equipment fail from wear. Usually, if such a period is known to exist within the useful life of the equipment, the equipment may be scheduled for overhaul before it fails from wear. Any item having a stipulated life which is substantially shorter than the life of the equipment should be considered "worn out" at the time it accomplishes its stipulated life. If it fails before its stipulated life, it should be counted as a relevant failure. If it is replaced before its stipulated life without its failing, the replacement should not be counted as a relevant failure. If it fails after its stipulated life, it should be regarded as a non-relevant failure.

(2) To calculate the lower percent confidence limit on the MTBF point estimate, the following should be used:

$$MTBF_{\mathrm{LCL}} = \frac{2\mathrm{T}}{\chi^2_{(\gamma,2F+2)}} \tag{A-5}$$

Where:

T = total test life (time, miles, rounds, cycles, etc.)
F = Total number of reliability failures
γ = confidence coefficient (such as 80%)

The denominator of Equation A-5 is determined using Chi-Squared (χ^2) tables or readily available computer programs or common spreadsheet software functions.

The lower percent confidence limit gives a statistical representation of the results that could be expected if the test was repeated or if a like item was subjected to the same conditions (i.e. a conservative estimate of test item performance). A graphical representation of a χ^2 distribution is displayed in Figure A-2:

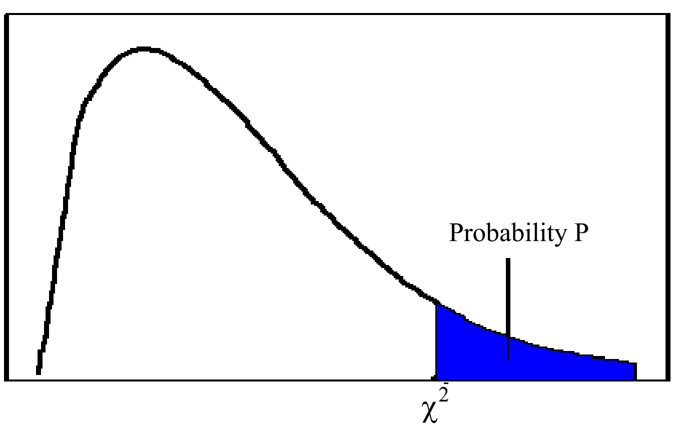

Figure A-2: χ^2 Distribution

The part of the distribution that we are interested in is the area to the left of the χ^2 value on the horizontal axis. There are many different iterations of this shape that depend on the degrees of freedom and the confidence level. We are interested in the area to the left of the χ^2 value because the customer is interested in a conservative estimate of how many miles will be traveled between failures. For example, a typical reliability criterion statement may read: Verify with 80% confidence that the MMBF will exceed 180 miles.

The confidence level is given in the customer requirement and the degrees of freedom are defined by 2F+2, where F is the total number of reliability failures.

Steps 4 and 5 in a reliability analysis both involve determining if there are differences between sets of data. Often, the evaluator is required to compare the results of multiple items that are the same model or to compare the results from one test with the results from a similar test. When this is the case, a statistical test should be performed to determine if there are statistically significant differences between the data sets. For the exponential distribution, a common method used toward this end, and the method discussed in the following sections, is the Chi-square (χ^2) Test for Independence. This test evaluates the equality of means to determine if the individual items are from the same statistical population. While the Chi-square Test is commonly used, there are more accurate methods and the Chi-square Test is not considered to be adequate when the number of failures per test item is less than five; therefore, the evaluator should determine the appropriate method to use each time a comparison is necessary.

The methodology for the Chi-square (χ^2) Test for Independence follows:

The hypothesis of principal interest is that the MTBF, or failure rates, are equal. The χ^2 Test for Independence compares a χ^2 value calculated from the data to a χ^2 critical value retrieved from a chi-square distribution table. If the calculated value is less than the critical value, the hypothesis is accepted.

The formula for the calculated χ^2 test statistic is:

$$\chi^2_{calc} = \sum \frac{(O_i - E_i)^2}{E_i} \qquad \text{(A-6)}$$

Where, O_i is the observed value and E_i is the expected value.

Once the calculated χ^2 value has been determined, the critical χ^2 value must be obtained. In order to look up the value in the χ^2 table, the degrees of freedom (dof) of the data and the acceptable level of significance (α) must be known.

The degrees of freedom of the data are simply the number of test items minus 1.

The level of significance is more difficult to understand. The level of significance is a kind of probability of error in a statistical test. It allows inferences to be made about the population using the data from the sample(s). The significance level is chosen by the test evaluator as an acceptable limit based upon several factors. The three main factors are the sample size, the magnitude of the observed values, and the degree of importance in what the data tells you (ex: chance of a water gun leaking vs. chance of a gun misfiring in a life or death situation).

First, as sample size is increased, confidence that the sample is representative of the entire population increases, so the expected probability of error (significance level) can be set lower. Second, if the magnitudes of the observed values increase while the size of the differences between the observed values remains the same, a lower significance level can be used. Third, the higher the degree of importance in what the data tells you, the lower the probability of error you are willing to accept.

Typically, developmental tests yield small sample sizes, observed values with small magnitudes, but at least a medium level of importance in what the data tells us because we are analyzing failures in Army materiel. Since a 5% significance level is frequently used when applying the χ^2 Test for Independence to much larger sample sizes with larger magnitudes of observed values but with less importance than our data, we have chosen a 10% significance level as the typical level for our calculations.

(3) Important Notes on Reliability Analysis:

(a) For all reliability analyses where the exponential assumption is used, the following statement should be included in the analysis: It is assumed that the time between failures is an exponentially distributed random variable. (Note: time would be changed to distance, rounds, etc. depending on the meters used for the calculations.)

(b) For all reliability analyses where multiple items of the same model are tested, the following statement should be included in the analysis: A statistical test was performed to determine if the reliability characteristics of same-model test items could be combined. (Note: If differences were observed at the prescribed significance level, the results could still be combined and presented for informational purposes as long as the results are preceded by a statement explaining that statistical differences were present.)

(c) When combination of the reliability results for multiple items is statistically acceptable, it is important to remember that this does not represent the results expected for a single item that was tested to the combined test life. Rather, combination of the results can be thought of as increasing the number of experiments run (i.e. sample size), which yields higher confidence in the results.

(d) The assumption of constant failure rate does not mean that failure modes are constant throughout the “usable” life of the item. Different types of failures will occur during different periods in the life of an item, but the overall failure rate will remain approximately constant.

b. EXAMPLES.

Provided herein are six examples which can be used in presenting the results of a reliability test.

(1) Reporting Results for Multiple Items When Combination of the Results is Acceptable and When Both the Point Estimate and the Lower 80% Confidence Limit Meet the Criterion.

Example Scenario: Four items are tested for 20,000 miles each. Test items 1-4 have 18, 23, 15, and 27 failures, respectively.

Requirement: The MMBF shall be no lower than 700 at the 80% LCL and the items shall achieve a reliability of at least 0.9 at the lower 80% confidence limit for completing a 75-mile mission.

Calculations/Analysis/Discussion:

(a) Calculate the MMBF point estimate, where

Point Estimate (PE) = Total number of test miles divided by the number of failures.

For Test Item 1: $MMBF_{PE} = 20{,}000 / 18 = 1111.1$

2: $MMBF_{PE} = 20{,}000 / 23 = 869.5$

3: $MMBF_{PE} = 20{,}000 / 15 = 1333.3$

4: $MMBF_{PE} = 20{,}000 / 27 = 740.7$

(b) Calculate MMBF lower 80% confidence limit [Using Equation (A-5)]:

For Test Item 1: $$MMBF_{LCL} = \frac{2 * 20{,}000}{\chi^2_{(0.8,2*18+2)}} = 887.3$$

2: $$MMBF_{LCL} = \frac{2 * 20{,}000}{\chi^2_{(0.8,2*23+2)}} = 714.3$$

3: $$MMBF_{LCL} = \frac{2 * 20{,}000}{\chi^2_{(0.8,2*15+2)}} = 1{,}039.8$$

4: $$MMBF_{LCL} = \frac{2 * 20{,}000}{\chi^2_{(0.8,2*27+2)}} = 618.6$$

(c) Calculate the mission reliability point estimate [Using Equation (A-2)]:

$$R(t) = e^{-\lambda t} = e^{-\frac{t}{MMBF_{PE}}}$$

Where, t is the mission duration.

$$\text{For Test Item 1:} \quad R(t) = e^{-\frac{75}{1111.1}} = 0.9347$$

$$2: \quad R(t) = e^{-\frac{75}{869.5}} = 0.9174$$

$$3: \quad R(t) = e^{-\frac{75}{1333.3}} = 0.9453$$

$$4: \quad R(t) = e^{-\frac{75}{740.7}} = 0.9037$$

(d) Calculate the mission reliability lower 80% confidence limit [Using Equation (A-2)]:

$$R(t) = e^{-\lambda t} = e^{-\frac{t}{MMBF_{LCL}}}$$

$$\text{For Test Item 1:} \quad R(t) = e^{-\frac{75}{887.3}} = 0.9324$$

$$2: \quad R(t) = e^{-\frac{75}{714.3}} = 0.9151$$

$$3: \quad R(t) = e^{-\frac{75}{1039.8}} = 0.9429$$

$$4: \quad R(t) = e^{-\frac{75}{618.6}} = 0.9014$$

Combination of Results:

In the above results, the observed values are the number of failures for each item. The expected values are derived from the assumption that the MMBF values are equal. This "expected" MMBF is the cumulative MMBF.

(e) Calculate the cumulative or "expected" MMBF:

$$\text{MMBF}_{\text{expected}} = \frac{80{,}000 \text{ miles}}{83 \text{ failures}} = 963.8$$

Likewise, the cumulative mission reliability can be calculated as follows:

$$R(t) = e^{-\frac{75}{963.8}} = 0.9251$$

By dividing each item's test life by 963.8 MMBF, we end up with a theoretical "expected" number of failures for each item. Note: Since the test life of all four items was equal in this example, the expected number of failures was equal. If the test lifes differ, the expected number of failures must be calculated for each separate item.

(d) Calculate the "expected" number of failures for each item:

$$\text{F}_{\text{expected}} = \frac{20{,}000 \text{ miles}}{963.8 \text{ miles}/\text{failure}} = 20.751 \text{ failures}$$

Notice that we end up with "partial" failures. This is okay because these "expected" values are only used in the χ^2 Goodness-of-Fit Test.

Next, we calculate the χ^2 terms for each vehicle using Equation (A-6) and then sum them to get the cumulative χ^2 value.

(e) Calculate the individual and cumulative χ^2 terms:

For Test Item 1: $$\chi^2_{\text{calc}} = \frac{(O_i - E_i)^2}{E_i} = \frac{(18 - 20.751)^2}{20.751} = 0.364$$

2: $$\chi^2_{\text{calc}} = \frac{(23 - 20.751)^2}{20.751} = 0.244$$

3: $$\chi^2_{\text{calc}} = \frac{(15 - 20.751)^2}{20.751} = 1.593$$

4: $$\chi^2_{\text{calc}} = \frac{(27 - 20.751)^2}{20.751} = 1.882$$

Cumulative χ^2: $0.364 + 0.244 + 1.593 + 1.882 = 4.084$

Once we have the cumulative χ^2 value, all we need is a critical value from the χ^2 table for comparison. In order to look up the value in the χ^2 table, we need to know the degrees of freedom (dof) of our data and the level of significance (α).

The degrees of freedom of our data are simply the number of test items minus 1. So in our example, dof = 3.

For our example, we will use a 10% level of significance (α).

All that is left is to lookup the χ^2 critical (threshold) value to compare with our cumulative χ^2 value. This can be done using readily available chi-square distribution tables or programs/spreadsheets with built in "solvers."

In our example,

$$\chi^2_{(0.1,\,3)} = 6.251394$$

We now have both of the χ^2 values needed to determine if we can combine our data. Before comparing these two values, we must understand what we are looking for in the comparison. A significance level is used to determine if differences in observed data are statistically significant. The χ^2 critical value is a lower limit we are willing to accept for significance. That is, if the calculated χ^2 value is greater than the χ^2 critical value, then the differences are said to be significant. Since, in our case, we would like to combine the observed data, we do not want any significant differences. Therefore, we want our calculated χ^2 value to be less than the χ^2 critical value. As a check for this, we can use a comparison ratio. The comparison ratio is simply the calculated χ^2 value divided by the χ^2 critical value. As long as this ratio is less than 1, we can combine the vehicles.

$$\text{Comparison Ratio} = 4.084 / 6.251 = 0.653$$

Conclusion: The combination of the test results for these four items is not rejected.

Suggestion for writing the above in report form: A statistical test for equality of means was performed to determine if combination of the test results for same-model items was statistically acceptable. No significant difference was present at the 10% significance level; therefore, the results were combined. The MMBF point estimate and the lower 80% confidence limit of the test items combined were 963.8 and 873.3, respectively. The MMBF criterion was met since the lower 80% confidence limit (873.3) was greater than the MMBF requirement of 700. The mission reliability point estimate and lower 80% confidence limit of the test items combined were 0.9251 and 0.9246, respectively. The reliability criterion was met since the lower 80% confidence limit (0.9246) was greater than the reliability requirement of 0.9 for a 75-mile mission. It was assumed that the distance between failures was an exponentially distributed random variable.

(2) Reporting Results for Multiple Items When Combination of the Results is Acceptable and When the Point Estimate Meets the Criterion But the Lower 80% Confidence Limit Does Not Meet the Criterion.

Example Scenario: Four vehicles are tested for 10,000 miles each. There are a combined total of 8 failures.

Requirement: The MMBF shall be no lower than 4,000 with a LCL of 80%.

Analysis/Discussion: Point Estimate = Total number of test miles divided by the number of failures = 40,000 / 8 = 5,000.

Lower 80% Confidence Limit = 3,515.

Suggestion for writing the above in report form: A statistical test for equality of means was performed to determine if combination of the test results for same-model items was statistically acceptable. No significant difference was present at the 10% significance level; therefore, the results were combined. The combined MTBF point estimate and the lower 80% confidence limit of the vehicles were 5,000 and 3,515, respectively. The criterion is not met since the lower 80% confidence limit (3,515) is less than the requirement of 4,000 miles. It is assumed that the distance between failures is an exponentially distributed random variable.

Additional notes/comments: Confidence limits are driven by sample size. In this particular example, if the sample size was increased to 10 items completing 10,000 miles each and the point estimate remained constant, the lower 80% confidence limit would be greater than the requirement; therefore, the criteria would have been met. By increasing the sample size when the point estimate meets the criteria and the LCL does not meet the criteria, the LCL can generally meet the criteria if the point estimate remains constant. However, in the testing community, one does not typically have the time or resources to increase the sample size to meet the required LCL. In some cases, this sample size would need to be extremely large. For additional information, the LCL will approach the point estimate as the sample size approaches infinity.

(3) Reporting Results for Multiple Items When Combination of the Results is Acceptable and When neither the Point Estimate nor the Lower 80% Confidence Limit Meet the Criterion.

Example Scenario: Two vehicles are tested for 20,000 miles each. One of the vehicles has five failures and the other vehicle has six failures.

Requirement: The MMBF shall be no lower than 10,000 with a LCL of 80%.

Analysis/Discussion: Point Estimate = Total number of test miles divided by the number of failures = 40,000 / 11 = 3,636.

Lower 80% Confidence Limit = 2,707.

Suggestion for writing the above example in report form: The MTBF point estimate and the lower 80% confidence limit of the vehicle were 3,636 and 2,707, respectively. The criterion is not met since both the point estimate and the lower 80% confidence limit are less than the MTBF requirement of 10,000. It is assumed that the number of miles between failures is an exponentially distributed random variable.

Additional notes/comments: The point estimate will always be greater than the lower 80% confidence limit (or any reasonable confidence limit). Therefore, if the point estimate is not greater than the requirement, the criterion is not met and the following statement could be written: The criterion is not met since the point estimate (3,636) is less than the requirement. (This is the one exception when the evaluator could use the point estimate to determine if the criteria are met. Never state that the criterion is met when only the point estimate is greater than the requirement, but it can be stated that the criterion is not met when the point estimate is less than the requirement.)

(4) Reporting Results for Multiple Items When Combination of the Results is Acceptable and When the Point Estimate and the Lower 90% Confidence Limit of the Combined Results Meets the Criterion but the Point Estimate and Lower 90% Confidence Limits of Some, or Even All, of the Individual Items does not Meet the Criterion.

Scenario: Ten weapons were tested for 36,000 rounds each. Each weapon incurred 15 failures (stoppages).

Requirement: The weapon shall have, as a minimum, mean rounds between stoppages (MRBS) of 2,100 rounds with 90% confidence.

Analysis/Discussion: The MRBS point estimates and lower 90% confidence limits of each individual weapon and combined are presented in Table A-1.

TABLE A-1. RELIABILITY CHARACTERISTICS

Item No.	Test Life (rounds)	Stoppages	MRBS Point Estimate	MRBS 90% LCL
1	36,000	15	2,400	1,690
2	36,000	15	2,400	1,690
3	36,000	15	2,400	1,690
4	36,000	15	2,400	1,690
5	36,000	15	2,400	1,690
6	36,000	15	2,400	1,690
7	36,000	15	2,400	1,690
8	36,000	15	2,400	1,690
9	36,000	15	2,400	1,690
10	36,000	15	2,400	1,690
Combined	360,000	150	2,400	**2,156**

Important Notes: This example illustrates the fact that confidence limits are driven by sample size. By increasing the sample size when the point estimate meets the criteria and the LCL does not meet the criteria, the LCL can generally meet the criteria if the point estimate remains constant. For example, had the sample size been half as large (5 weapons firing 36,000 rounds each) with the same number of stoppages per weapon (15), the combined MRBS 90% LCL would have been 2,060 rounds and the criterion would not have been met. However, in the testing community, quite often the time or resources required to increase the sample size to meet the LCL are not available. In some cases, this sample size would need to be extremely large.

Suggestion for writing the above example in report form: The MRBS point estimate and the lower 90% confidence limit of all ten weapons combined were 2,400 and 2,156 rounds, respectively. The criterion is met since the lower 90% confidence limit is greater than the requirement of 2,100 rounds. It is assumed that the number of rounds between failures is an exponentially distributed random variable.

(5) Point Estimate Not Meeting the Criterion but the Lower 80% Confidence Limit does Meet the Criterion.

This example is not possible. The point estimate will always be greater than the lower 80% confidence limit or any other reasonable confidence limit (>80% LCL). Therefore, it is not possible for the lower 80% confidence limit, or any other reasonable confidence limit, to meet the criterion when the point estimate does not meet the criterion.

(6) Less than the Minimum Test Life.

Example Scenario: One item was tested for 3,000 miles without incurring any failures.

Requirement: The MMBF shall be no lower than 3,000 with a LCL of 80%.
Analysis/Discussion: Point Estimate = total number of test miles / total number of reliability failures = 3,000 / 0 = Undefined (Divided by 0).

MMBF Lower 80% Confidence Limit = 1,864.

This is an example of a “statistically inadequate” test for the following reason: There is no mathematical/statistical way that a test item could meet this requirement with a test life of 3,000 miles. This requirement requires a minimum of 2 items completing 3,000 miles without any failures. A good rule-of-thumb to follow when designing a reliability test is to plan the test for a minimum of 2.5 times the reliability criterion.

Suggestion for writing the above example in report form: The MMBF point estimate could not be calculated since no failures were incurred. The lower 80% confidence limit of the test item was 1,864 miles. Since the lower 80% confidence limit is less than the requirement, the criterion is not met. It should also be noted that this requirement was impossible to meet with a single test item completing 3,000 miles (statistically inadequate test). In order to meet this requirement with 80% confidence, 2 items would need to complete 3,000 miles each without a failure. It is assumed that the number of miles between failures is an exponentially distributed random variable.

APPENDIX B. DURABILITY OR MISSION RELIABILITY ANALYSIS USING THE BINOMIAL DISTRIBUTION.

a. ANALYSIS DISCUSSION.

Durability testing is generally carried out by operating a group of test items under conditions that duplicate field conditions as closely as possible until durability failures occur or some predetermined test time occurs. The predetermined required mission time should be designed to be some percent (minimum of 20% preferred) of the anticipated service life of the item, agreed upon by the developer and the testing activity/evaluator.

Point estimates and the LCLs will be calculated. The point estimate is the quotient of the number of items that successfully completed the required mission and the total number of items. The LCL (90% preferred) is determined by the application of the binomial distribution (See Equation (B-1)).

The criteria to be assessed should clearly state whether the point estimate or the LCL will be evaluated. If the criteria are unclear, then request clarification from the Program Manager or the evaluator. The point estimate is the arithmetic mean of the time between failures and equates to a very low confidence. The LCL is the best estimate of the true unknown success rate. Upper confidence limits are not generally calculated since the evaluator is typically not interested in how “good can the product be,” but rather in what is the “worst case” that can be expected.

Methodology.

The steps required to address a durability or mission reliability criterion through application of the binomial distribution are:

(1) Assemble all durability failure or reliability failure data (incident scores, chargeabilities, times, etc.)

(2) Verify that the binomial distribution is appropriate for the data set being analyzed.

(3) Calculate the durability or mission reliability parameters.

(4) Determine if the results for multiple test items can be combined.

(5) Look for statistical differences between similar items, other tests, subsystems, etc.

(6) Assess the criterion.

The following contains detailed explanations of Steps 2 through 5, followed by thirteen examples which cover some typical durability or mission reliability scenarios that can be used in assessing the criteria.

Step 2 is verifying that the binomial distribution is appropriate for the data set. The binomial distribution describes the behavior of a count variable if the following conditions apply:

(1) The number of observations is fixed.

(2) Each observation is independent.

(3) Each observation represents one of two outcomes ("success" or "failure").

(4) The probability of "success" is the same for each outcome.

If any one of these conditions does not apply, the assumption of the binomial distribution is invalid. When the distribution is unknown, statistical "goodness-of-fit" tests (e.g. chi-squared, Kolmogorov-Smirnov, Cramer-Smirnov-Von-Mises, and Anderson-Darling) may be used to determine whether the assumption of a particular distribution is reasonable.

Step 3 is calculating the durability or mission reliability parameters. The parameters should always be presented as a point estimate coupled with either a confidence limit or a confidence interval to qualify the results. The confidence level gives a statistical representation of the results that could be expected if the test were repeated. The accuracy of the confidence level, or interval, depends on the quality and quantity of the available data. One of the most common statistical distributions used for calculating durability parameters, and the distribution covered in the following sections, is the binomial distribution.

Binomial Distribution:

(1) The point estimate is evaluated as the quotient of the number of items that successfully completed the required mission and the total number of items.

(2) The LCL (γ) is calculated using the binomial probability mass function:

$$\sum_{x}^{n} \binom{n}{x} p^{x}(1-p)^{n-x} = 1-\gamma \tag{B-1}$$

Where:

n = number of items tested
x = number of successes
p = probability of success
γ = Lower confidence limit
$\binom{n}{x} = \frac{n!}{x!(n-x)!}$

To find γ for a specific confidence (such as 80% or 90%), an iterative procedure would need to be used. This iterative procedure can be done using various spreadsheet "solvers" or software packages. Also, binomial reliability tables and computer programs are readily available for determining reliability estimates. (In all cases, verify the results of these packages "manually" before using them.)

There are "shortcut" formulas for determining probability of success, LCL, and sample size when the number of failures (durability failures) is exactly 0. Equations B-2, B-3, and B-4 are derived from Equation B-1 by setting x (number of successes) equal to n (number of items tested) and can be used only when the number of failures is exactly 0.

$$p = (1-\gamma)^{1/n} \tag{B-2}$$

Note that the above formula can be written in the following forms to find confidence limits and sample sizes:

$$\gamma = 1 - p^n \tag{B-3}$$

$$n = \ln(1-\gamma)/\ln(p) \tag{B-4}$$

For all durability analyses, the following statement should be included in the analysis: It is assumed that the number of successes per trials is a binomially distributed random variable.

b. <u>EXAMPLES.</u>

Provided herein are thirteen examples which can be used in design of a durability test and in presenting the results. Some of the examples contain "reasonably" large sample sizes with respect to the testing community. These "large sample sizes" are being utilized to show whether requirements are met/not met under various unique situations. In the testing arena, the sample size is generally small (less than 10). It will be shown in the examples with a "small" sample size that it can be difficult to meet some requirements with a reasonable confidence. This is something that must be considered in developing the test plan. If it is known that either the test time (miles, rounds, hours, cycles, etc) or the sample size is not large enough to statistically demonstrate the requirement, the following statement should be included in the test plan: "It will not be possible to statistically meet the criteria due to the small sample size or the limited amount of test time." This will inform the evaluator and materiel developer that the test is a statistically inadequate test up front.

(1) Determining Confidence Level at Which Requirement is Demonstrated [Using Formula (B-1)].

This example problem shows how to determine γ when the probability of a success (p), number of items tested (n), and number of successes (x) are given.

Given: p = 0.8, n = 25, and x = 22

Required: γ

Solution:

$$\sum_{x=22}^{n=25} \binom{n}{x} 0.8^{x} (1-0.8)^{n-x} = 1-\gamma$$

$$\binom{25}{22} 0.8^{22}(1-0.8)^{3} + \binom{25}{23} 0.8^{23}(1-0.8)^{2} + \binom{25}{24} 0.8^{24}(1-0.8)^{1} + \binom{25}{25} 0.8^{25}(1-0.8)^{0} = 1-\gamma$$

2300 * 0.007379 * 0.008 + 300 * 0.005903 * 0.04 + 25 * 0.004722 * 0.2 + 1 * 0.003778 * 1 =

0.135768 + 0.070835 + 0.023612 + 0.003778 =

0.233993 = 1 – γ

γ = 0.766007 = 76.6007%

This solution can be stated as follows: It can be stated with 76.6% confidence that the probability is 0.8 of completing a successful mission when there are 22 successes in 25 trials (i.e. 3 of the 25 test items had durability failures). It is assumed that the number of successes per trials is a binomially distributed random variable. An iterative routine could be utilized with the previous example by defining γ to equal 0.9 and then solving for p (probability of success). The solution could then be stated as follows: It can be stated with 90% confidence that the probability is 0.752 of completing a successful mission when there are 22 successes in 25 trials. (This is the more conventional method used in the test arena.)

(2) Determining Probability of Success (Durability) at Fixed Lower Confidence Level (Using "Shortcut" Equation B-2).

Thirty of thirty items completed a required mission without a durability failure. Determine the probability of success when the LCL is 0.9.

$$p = (1-\gamma)^{1/n} = (1-0.9)^{1/30} = 0.926119$$

Therefore, it can be stated with at least 90% confidence that the probability of completing a mission when 30 of 30 items complete a mission without a durability failure is equal to 0.926. It is assumed that the number of successes per trials is a binomially distributed random variable.

(3) Determining Lower Confidence Level for Meeting a Fixed (Known) Mission (Durability) (Using "Shortcut" Equation B-3).

Thirty of thirty items completed a required mission without a durability failure. Determine the LCL when the probability of a success is 0.9.

$$\gamma = 1 - p^n = 1 - 0.9^{30} = 1 - 0.042391 = 0.957609$$

Therefore, it can be stated with at least 95.7% confidence that the probability of completing a mission when 30 of 30 items complete a mission without a durability failure is equal to 0.9. It is assumed that the number of successes per trials is a binomially distributed random variable.

(4) Determining Sample Size to Demonstrate a Specified Mission (Durability) Success as a Specified Confidence (Using "Shortcut" Equation B-4).

Determine the number of items required in order to have a probability of success of 0.95 and a LCL of 0.90.

$$n = \ln(1-\gamma)/\ln(p) = \ln(1-0.90)/\ln(0.95) = 44.89057 = 45$$

To achieve a probability of success of 0.95 and a LCL of 0.90, 45 of 45 test items would need to complete the mission without a failure (durability failure). Therefore, it can be stated with at least 90% confidence that the probability of completing a mission when 45 of 45 items complete a mission without a durability failure is equal to 0.95. It is assumed that the number of successes per trials is a binomially distributed random variable.

The following examples are shown using various "real life" durability situations. All of the examples use mission time (predetermined test time) in hours. This type of procedure can also be utilized when the mission time is in days, cycles, rounds, miles, kilometers, erect/strike tents, don/doff clothed, etc. The lower 90% confidence limit is used to determine if the item met the criteria unless the requirement specifies a confidence as shown in Example 13. As stated earlier, a lower confidence less than 80% should not be considered unless specified in the requirement. If so, all concerned parties should be aware of the implication of such a low confidence limit.

(5) Reporting Results When the Point Estimate Meets the Criterion But the Lower 90% Confidence Limit Does Not.

Example Scenario: Ten items are tested to a required mission life of 100 hours. Eight of the ten items completed the required mission life of 100 hours without a durability failure.
Requirement: Probability of completing a successful mission of 100 hours is 0.70.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 100 hours divided by the number of items which attempted the test (number of trials) = 8 / 10 = 0.800.

Lower 90% Confidence Limit is calculated to be 0.55039 based on equation B-1.
Suggestion for writing the above in report form: The point estimate and the lower 90% confidence limit of the test item were 0.800 and 0.550, respectively. The criterion is not met since the lower 90% confidence limit (0.550) is less than the requirement of 0.70. It is assumed that the number of successes per trials is a binomially distributed random variable.

Additional notes/comments: Confidence limits are driven by sample size. In this particular example, if the sample size was increased to 50 and the point estimate remained constant, the lower 90% confidence limit would be greater than the requirement; therefore, the criteria would have been met. By increasing the sample size when the point estimate meets the criteria and the LCL does not meet the criteria, the LCL can generally meet the criteria if the point estimate remains constant. However, in the testing community, one does not typically have the time or resources to increase the sample size to meet the required LCL. In some cases, this sample size would need to be extremely large. For additional information, the LCL will approach the point estimate as the sample size approaches infinity.

(6) Reporting Results When neither the Point Estimate nor the Lower 90% Confidence Limit Meet the Criterion.

Example Scenario: Forty-five items are tested to a required mission life of 100 hours. Thirty of the forty-five items completed the required mission life of 100 hours without a durability failure.

Requirement: Probability of completing a successful mission of 100 hours is 0.85.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 100 hours divided by the number of items which attempted the test (number of trials) = 30 / 45 = 0.667.

Lower 90% Confidence Limit is calculated to be 0.56116 based on equation B-1.

Suggestion for writing the above example in report form: The point estimate and the lower 90% confidence limit of the test item were 0.667 and 0.561, respectively. The criterion is not met since both the point estimate and the lower 90% confidence limit are less than the requirement of 0.85. It is assumed that the number of successes per trials is a binomially distributed random variable.

Additional notes/comments: The point estimate will always be greater than the lower 90% confidence limit (or any reasonable confidence limit). Therefore, if the point estimate is not greater than the requirement, the criterion is not met and the following statement could be written: The criterion is not met since the point estimate (0.667) is less than the requirement. (This is the one exception when the evaluator could use the point estimate to determine if the criteria are met. Never state that the criterion is met when only the point estimate is greater than the requirement, but it can be stated that the criterion is not met when the point estimate is less than the requirement.)

(7) Reporting Results When the Point Estimate and Lower 90% Confidence Limit Both Meet the Criterion.

Scenario: Forty-five items are tested to a required mission life of 100 hours. Forty-three of the forty-five items completed the required mission life of 100 hours without a durability failure.

Requirement: Probability of completing a successful mission of 100 hours is 0.85.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 100 hours divided by the number of items which attempted the test (number of trials) = 43 / 45 = 0.956.

Lower 90% Confidence Limit is calculated to be 0.88602 based on equation B-1.

Suggestion for writing the above example in report form: The point estimate and the lower 90% confidence limit of the test item were 0.956 and 0.886, respectively. The criterion is met since the lower 90% confidence limit is greater than the requirement of 0.85. It is assumed that the number of successes per trials is a binomially distributed random variable.

(8) Point Estimate Not Meeting the Criterion but the Lower 90% Confidence Limit does Meet the Criterion.

This example is not possible. The point estimate will always be greater than the lower 90% confidence limit or any other reasonable confidence limit (>80% LCL). Therefore, it is not possible for the lower 90% confidence limit, or any other reasonable confidence limit, to meet the criterion when the point estimate does not meet the criterion.

(9) Reporting Results When Test Items are Repaired to “Like New” During Test.

Example Scenario: Ten items were tested to a required mission life of 100 hours. Four of the items had durability failures at 50, 60, 70, and 80 hours. The remaining six items completed the required mission of 100 hours. However, the items which had durability failures at 50 and 60 hours were completely rebuilt and then both items restarted the 100 hr test. One of these two “new” items completed the required mission life of 100 hours without a durability failure while the other one experienced a durability failure at 45 hours. (Note: In this example, it is assumed that the number of test items is 12 (10 original items and 2 “new” items.)

Requirement: Probability of completing a successful mission of 100 hours is 0.50.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 100 hours divided by the number of items which attempted the test (number of trials) = 7 (6 of the original and 1 of the new items) / 12 (10 original and 2 new items) = 0.583.

Lower 90% Confidence Limit is calculated to be 0.36228 based on equation B-1.

Suggestion for writing the above example in report form: The point estimate and the lower 90% confidence limit of the test item were 0.583 and 0.362, respectively. The criterion is not met since the lower 90% confidence limit (0.362) is less than the requirement of 0.85. It is assumed that the number of successes per trials is a binomially distributed random variable.

(10) Reporting Results When Item “Almost” Meets Requirement.

It should be noted that either an item successfully completes some predetermined time, or it does not. For example, an item would not successfully complete a mission life of 100 hours if it should have a durability failure at 99 hours. Therefore, that item would be considered a failure from the “Durability Definition.” Let’s look at the following example:

Example Scenario: All forty-five items had durability failures at times between 90 and 100 hours with a requirement of 100 hours without a durability failure.

Requirement: Probability of completing a successful mission of 100 hours is 0.50.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 100 hours divided by the number of items which attempted the test (number of trials) = 0 / 45 = 0.000.

Lower 90% Confidence Limit is calculated to be 0.000 based on equation B-1.

Suggestion for writing the above example in report form: The point estimate and the lower 90% confidence limit of the test item were 0.0 and 0.0, respectively. Therefore, the criterion is not met. However, it should be noted that all 45 items completed at least 90 hours without a durability failure. If the requirement would have been 90 hours without a durability failure instead of 100 hours without a durability failure, the point estimate and the lower 90% confidence limit would have been 1.00 and 0.95, respectively. Therefore, the item does show potential. It is assumed that the number of successes per trials is a binomially distributed random variable.

(11) Items Having Durability Failures and Being Fixed “Like New” (Similar to Example 9).

Example Scenario: One-hundred items were tested to a required mission life of 100 hours. Nine of the items had durability failures at 30, 35, 40, 50, 60, 65, 80, 87, and 95 hours. All nine of these items were completely rebuilt and were tested again as “new” items. Five of these nine items had durability failures at 15, 20, 23, 38, and 41 hours (i.e. 4 of these 9 rebuilt items completed a mission of 100 hours without a durability failure).

Requirement: Probability of completing a successful mission of 100 hours is 0.85.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 100 hours divided by the number of items which attempted the test (number of trials) = 95 (91 of the original and 4 of the new items) / 109 (100 original and 9 new items) = 0.87156.

Lower 90% Confidence Limit is calculated to be 0.82059 based on equation B-1.

Suggestion for writing the above example in report form: The point estimate and the lower 90% confidence limit of the test item were 0.872 and 0.821, respectively. The criterion is not met since the lower 90% confidence limit (0.821) is not greater than the requirement of 0.85. It is assumed that the number of successes per trials is a binomially distributed random variable.

Additional Comments: This is an example of a situation in which the original test item would have met the requirement because of the following: The point estimate of the original test is equal to the quotient of 91 (number of items which completed a mission life of 100 hours) and 100 (number of items which attempted the original test). Therefore, the point estimate is equal to 0.91. The lower 90% confidence limit was 0.861. Since the lower 90% confidence limit was greater than the requirement of 0.85, the criterion would have been considered met. Therefore, it should be noted that the use of "new" items will not necessarily increase the point estimate and LCL. (The point estimate will never be increased by using "new" items when the point estimate of the "new" items is less than the point estimate of the original test items.)

(12) Reporting Results When Some of the Items Have Test Times Greater Than the Requirement (Enhancement of Sample Size).

Example Scenario: Two out of eight items had durability failures at 10 and 24 hours. The other six items were tested to the following times without a durability failure: 103, 75, 75, 70, 25, and 210 hours. (Note: this situation rarely occurs in the real world since very few items are tested beyond the minimum number of hours required for a successful mission. In this case, the minimum number of hours required for a successful mission is 25 hours as shown in the requirement.)

Requirement: Probability of completing a successful mission of 25 hours is 0.75.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 25 hours divided by the number of items which attempted the test (number of trials) = 21 / 23 = 0.913.

Note: the number of items which successfully completed a mission life of 25 hours without a durability failure was 21 (1 item completed 4 successful missions of 25 hours plus an additional 3 hours, 2 items completed 3 successful missions of 25 hours each for a total of 6 successful missions of 25 hours each, 1 item completed 2 successful missions of 25 hours plus an additional 20 hours, 1 item completed 1 mission of 25 hours, and 1 item completed 8 successful missions of 25 hours plus an additional 10 hours). The number of items which attempted the test was 23 items (1 + 1 + 4 + 3 + 3 + 2 + 1 + 8 = 23).

Lower 90% Confidence Limit is calculated to be 0.785 based on equation B-1.

Suggestion for writing the above example in report form: The point estimate and the lower 90% confidence limit of the test item were 0.913 and 0.785, respectively. The criterion is met since the lower 90% confidence limit is greater than the requirement of 0.75. It is assumed that the number of successes per trials is a binomially distributed random variable.

Additional Comments: This is an example of a situation in which five of the original test items were tested beyond the requirement. In this case, the item which was tested successfully for 103 hours is considered to be equivalent to four test items with test times of 25 hours without a durability failure. This is a conservative method of determining the number of test items. There has been work done which will show the number of successes to be somewhat larger and will possibly give a somewhat larger point estimate and LCL. However, this method is seldom used in the real world. Also, very few times in the real world does a test item incur more than the minimum number of required test miles.

(13) Less than the Minimum Number of Items and Extremely Low Confidence Limit.

Example Scenario: Two items were tested for 100 hours without a durability failure.

Requirement: Probability of completing a successful mission of 100 hours is 0.90 with a LCL of 50%.

Analysis/Discussion: Point Estimate = number of items which completed a mission life (successes) of 100 hours divided by the number of items which attempted the test (number of trials) = 2 / 2 = 1.00.

Lower 50% Confidence Limit is calculated to be 0.70 based on equation B-3.

Important Notes: This is an example of a "statistically inadequate" test for the following reason: There is no mathematical/statistical way that a test item could meet this requirement with a sample size of only 2. This requirement requires a minimum sample size of 7 without any failures (See Example 4). Also, it should be noted that 50% is an extremely low confidence limit. In short, a 50% confidence limit is not much better than a "coin-flip." When using this low confidence limit, it should always be stated that 50% is an extremely low confidence limit and state that as a minimum the lower 80% confidence should be used and give the results with 80% confidence, also.

Suggestion for writing the above example in report form: The point estimate, the lower 50% confidence limit, and the lower 80% confidence limit of the test item were 1.00, 0.70, and 0.44, respectively. Since the lower 50% confidence limit (an extremely low confidence limit) is less than the requirement, the criterion is not met statistically. It should also be noted that this requirement was impossible to meet with a sample size of 2 (statistically inadequate test). In order to meet this requirement with 50% confidence, 7 items would need to complete 100 hours without a durability failure. In order to meet this requirement with a more reasonable confidence, such as 80%, sixteen items would have to complete 100 hours without a durability failure. It is assumed that the number of successes per trials is a binomially distributed random variable.

APPENDIX C. ACRONYMS.

AMC	Army Materiel Command
BITE	Built-in Test Equipment
CFE	Contractor Furnished Equipment
CLS	Contractor Logistic Support
ECM	Electronic Control Module
ECU	Electronic Control Units
EFF	Essential Function Failures
EMA	Essential Maintenance Actions
EPR	Environmental Program Requirements
FD/SC	Failure Definition/Scoring Criteria
FPT	Follow-on Production Test
FSR	Field Service Representative
HEDGE	Human Engineering Data Guide for Evaluation
ILS	Integrated Logistic Supportability
LCL	Lower Confidence Limit
MAC	Maintenance Allocation Chart
MaxTTR	Maximum Time to Repair
MMBF	Mean Miles Between Failures
MR	Maintenance Ratio
MRBS	Mean Rounds Between Stoppages
MTBF	Mean Time Between Failures
MTTR	Mean Time to Repair
NEFF	Non-Essential Function Failures
NSN	National Stock Number
OMF	Operational Mission Failures
OMS/MP	Operational Mode Summary and Mission Profile
PE	Point Estimate
PM	Program Manager
PVT	Production Verification Test
RAM-D	Reliability, Availability, Maintainability, Durability
RPSTL	Repair Parts Special Tools List

SA	System Aborts
SAC	Supportability Analysis Chart
SAR	Safety Assessment Report
SMR	Source, Maintenance, and Recoverability
SOP	Standard Operating Procedures
SSC	Supply Support Chart
SSP	System Support Package
SSPCL	System Support Package Components List
STEC	Support and Test Equipment Chart
TDEP	Technical Data/Equipment Publications
TIR	Test Incident Report
TM	Technical Manual
TOP	Test Operations Procedure
TRADOC	US Army Training and Doctrine Command
UMA	Unscheduled Maintenance Actions

APPENDIX D. REFERENCES.

Required publications.

1. TOP 2-2-506, Endurance Testing of Tracked and Wheeled Vehicles, 26 June 1981.

2. DA PAM 73-1, Test and Evaluation in Support of System Acquisition, 30 May 2003

3. MIL-STD-38784, Standard Practice for Manuals, Technical: General Style and Format Requirements, 2 July 1995.

4. MIL-M-63000™ Series, Manuals, Technical.

5. SAE J1939, Recommended Practice for a Serial Control and Communications Vehicle Network, October 2007

6. SAE J1708, Serial Data Communications Between Microcomputer Systems in Heavy-Duty Vehicle Applications, August 2004

7. AMC Pamphlet 706-134, Engineering Design Handbook-Maintainability Guide for Design, 1972.

8. MIL-STD-1472F, Human Engineering Design Criteria for Military Systems, Equipment, and Facilities, 1998

For information only (related publications).

a. MTP 2-2-503, Maintenance (Vehicle), 5 June 1966.

b. MTP 3-1-002, Confidence Intervals and Sample Size, 25 January 1967.

c. TOP 7-3-507, Maintenance (Maintainability/Availability), 1 September 1983.

d. MIL-STD-721C, "Definition of Terms for Reliability and Maintainability", 12 June 1981.

e. MTP 10-2-507, Maintenance Evaluation, 15 September 1971.

f. AR 702-3, Army Materiel Systems Reliability, Availability, and Maintainability (RAM), 1 May 1982.

g. AR 70-10, Test and Evaluation during Development and Acquisition of Materiel.

h. ITOP 2-2-506(1), Tracked Vehicle Endurance Testing, 15, May 1984.

i. MTP 10-2-501, Operator Training and Familiarization, 27 March 1967.

j. MIL-HDBK-781A, Handbook for Reliability Test Methods, Plans, and Environments for Engineering Development, Qualification, and Production, 1 April 2006

Forward comments, recommended changes, or any pertinent data which may be of use in improving this publication to the following address: Test Business Management Division (TEDT-TMB), US Army Developmental Test Command, 314 Longs Corner Road Aberdeen Proving Ground, MD 21005-5055. Technical information may be obtained from the preparing activity: RAM/ILS Engineering and Analysis Division (TEDT-AT-ADR), US Army Aberdeen Test Center, Aberdeen Proving Ground, MD 21005-5059. Additional copies can be requested through the following website: http://itops.dtc.army.mil/RequestForDocuments.aspx, or through the Defense Technical Information Center, 8725 John J. Kingman Rd., STE 0944, Fort Belvoir, VA 22060-6218. This document is identified by the accession number (AD No.) printed on the first page.