DEVCOM DAC-TR-2023-018
March 2023

Reliability Growth Planning with Reliability Assurance Testing

by Martin Wayne

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position unless so specified by other official documentation.

WARNING

Information and data contained in this document are based on the input available at the time of preparation.

TRADE NAMES

The use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software. The report may not be cited for purposes of advertisement.

DEVCOM DAC-TR-2023-018
March 2023

Reliability Growth Planning with Reliability Assurance Testing

by Martin Wayne
DEVCOM Analysis Center

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	3. DATES COVERED END DATE
March 2023	Technical Report	1/3/22	9/30/22

4. TITLE AND SUBTITLE
Reliability Growth Planning with Reliability Assurance Testing

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER	5e. TASK NUMBER	5f. WORK UNIT NUMBER

6. AUTHOR(S)
Martin Wayne

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Director DEVCOM Analysis Center 6896 Mauchly Street Aberdeen Proving Ground, MD 21005	DEVCOM DAC-TR-2023-018

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Rapid fielding of new military systems has significantly reduced acquisition program schedules and led to reductions in traditional reliability growth test programs. Bayesian statistical techniques can be helpful in these environments, as they provide a more complete assessment of reliability through the combination of disparate data sources (e.g., different tests or configurations). The model results may also be used to develop assurance test plans, which are the Bayesian analogue to traditional reliability demonstration tests. This report outlines an approach for determining the amount of testing necessary for reliability assurance at the end of a test program, while also aligning these results with a reliability growth test program prior to the assurance test. The approach provides a reasonable path for growing reliability within the constrained environment.

15. SUBJECT TERMS
reliability growth, Bayesian assurance testing, reliability demonstration testing, fix effectiveness factors

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	29

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Martin Wayne	(410) 278-3221

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

(U) Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

The increased desire to rapidly field new military systems has resulted in significant reductions to acquisition program schedules. These reductions present numerous problems for ensuring that new systems have sufficient reliability. Many past acquisition programs have chosen to rely heavily on reliability growth during developmental testing to improve the overall system reliability and meet requirements. The expense of full system tests for some programs may also make it difficult to obtain meaningful testing regardless of the desired schedule. Bayesian statistical techniques are one option that may be beneficial in such a constrained environment. They provide a straightforward method for combining data from disparate sources (e.g., different tests or configurations), which may provide a more complete assessment of reliability than examining each data source independently. The model results may also be used to develop assurance test plans, which are the Bayesian analogue to traditional reliability demonstration tests (RDTs) in which prior information is used along with the test results. This report outlines an approach for determining the amount of testing necessary for reliability assurance at the end of a test program, while also aligning these results with a reliability growth test program directly preceding the assurance test. The approach provides a reasonable path for growing reliability to a level that will meet the desired risks of passing the assurance test.

2. TRADITIONAL RELIABILITY GROWTH PLANNING

Reliability growth planning is centered on the development of a reliability growth planning curve that depicts how reliability is expected to grow throughout a developmental test program. The curve reflects the discovery and subsequent corrective actions of failure modes during the developmental testing. Development of the planned growth curve generally relies on experience from previous programs to model the growth that can be expected. Reliability data collected during testing are then compared with the planned values to determine if the reliability is improving in a satisfactory manner.

2.1 Determining the End Point

The initial step in developing the reliability growth planning curve is the determination of the end point for the curve. This process involves an RDT, and first considers the reliability requirement along with any statistical confidence levels that are required. The parameters defining the RDT consist of the test duration T and the allowable number of failures c. The "acceptance" or "passing" criterion for the test is observing no more than c failures.

The number of allowable failures is found by bounding the probability of passing the test at the desired significance level (i.e., 1 – confidence) when the true Mean Time Between Failure (MTBF) of the system is equal to the MTBF requirement. The cumulative Poisson distribution is used, and c is the largest integer value of k satisfying the inequality in Eq. 1:

$$\sum_{i=0}^{k} \frac{\left(\frac{T}{M_R}\right)^i}{i!} exp\left(-\frac{T}{M_R}\right) \leq 1-\alpha. \quad (1)$$

For the inequality in Eq. 1, α is the significance, T is the length of demonstration test, and M_R is the MTBF requirement.

The probability of passing the demonstration test is then just the probability of observing at most the allowable number of failures. This probability is a function of the unknown MTBF, M. It again uses the Poisson distribution as in Eq. 2, where n is the observed number of failures.

$$p(n \leq c) = \sum_{i=0}^{c} \frac{\left(\frac{T}{M}\right)^i}{i!} exp\left(-\frac{T}{M}\right) \quad (2)$$

Plotting the probability in Eq. 2 as a function of M yields the Operating Characteristic (OC) curve for the demonstration test. The curve is specific to the allowable number of failures and demonstration test length, which also makes it an implicit function of the desired confidence level for the demonstration. An example OC curve is shown in Figure 1.

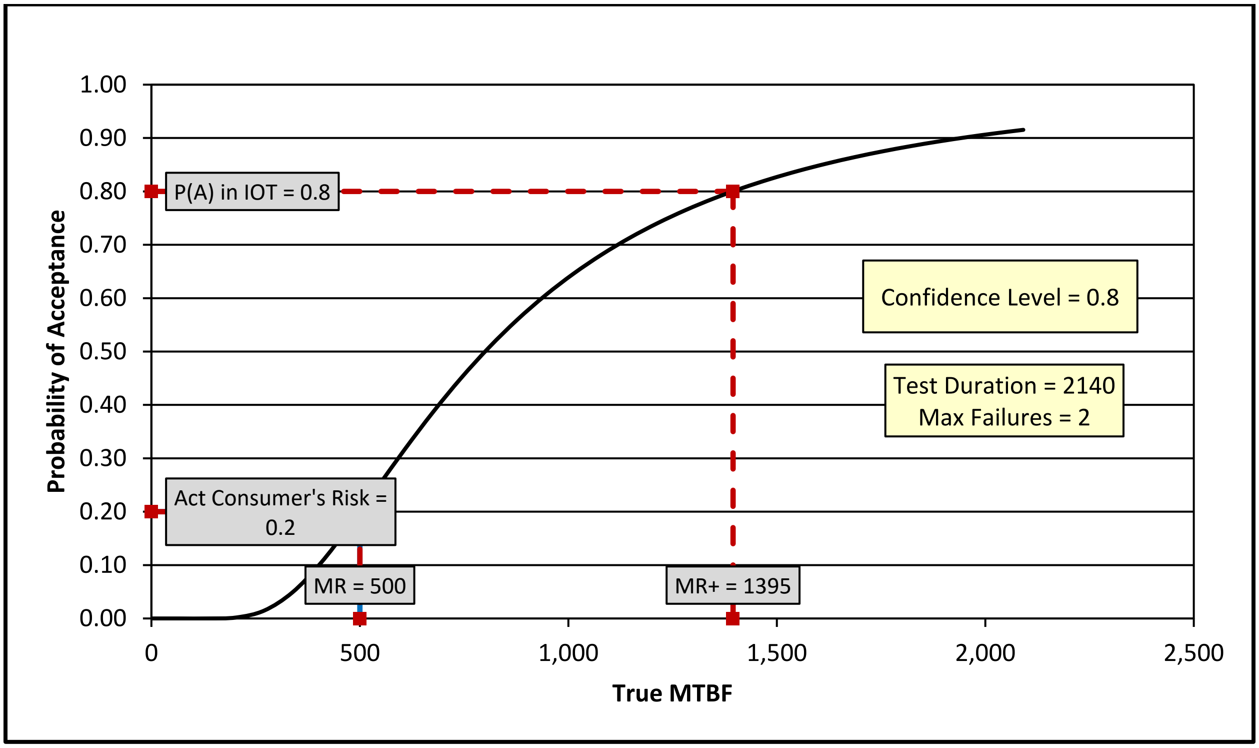

Figure 1. Example OC curve for RDT

2.2 Determining the Starting Point

The initial reliability for a system under development is not typically known when the planned curve is developed. A starting point for the planned growth curve may be determined by specifying a minimum level of reliability (or maximum failure rate) that must be achieved in early testing to have assurance that the reliability goals will be met through the planned reliability growth process.

Two important management metrics are necessary to determine the minimum initial reliability. The first is the average Fix Effectiveness Factor (FEF), which is commonly written as μ_d. The FEF is the fractional reduction in the failure rate for an observed failure mode after a corrective action has been implemented. The average FEF used for reliability growth planning represents the average level of fix-effectiveness that is planned across all observed failure modes. The second metric is the Management Strategy (MS), which is the proportion of the initial failure rate comprising failure modes that will be corrected when observed. It is directly proportional to the amount of reliability growth that can be achieved during a test program, and it is further defined as

$$MS = \frac{\lambda_B}{\lambda_I} \,. \tag{3}$$

λ_B is the total initial failure rate for failure modes that will have corrective actions implemented when observed, and λ_I is the total initial failure rate.

The key concept in finding the minimum initial reliability is the reliability growth potential. The growth potential failure rate, λ_{GP}, is the theoretical lower bound on the failure rate that can be achieved after all modes are observed and corrected with an assumed average level of fix effectiveness. It can be expressed as

$$\lambda_{GP} = \lambda_I(1 - \mu_d MS), \tag{4}$$

and is often used to manage risk when planning a test program. The growth potential itself is generally not achievable in the limited testing associated with a reliability growth program, which means the growth potential failure rate needs to be lower than the desired goal failure rate for the growth program. Recommendations for the ratio of the goal-to-growth potential are based on the MTBF (or reciprocal of the failure rate). The recommended ratio of goal MTBF to growth potential MTBF is to be at most 0.8, with lower risk being associated with lower values. The ratio can be defined using the goal and growth potential failure rates as

$$R = \frac{\lambda_{GP}}{\lambda_G}. \tag{5}$$

Combining Eqs. 4 and 5 then yields

$$\lambda_G = \frac{\lambda_I(1-\mu_d MS)}{R} \,. \tag{6}$$

The expression in Eq. 6 can then be rewritten as

$$\lambda_I = \frac{\lambda_G R}{(1-\mu_d MS)}, \tag{7}$$

which is a direct expression of the maximum initial failure rate (minimum initial MTBF) that is required to achieve the desired end point on the planning curve.

2.3 Constructing the Planning Curve

There are several specific reliability growth planning models, but the Planning Model Based on Projection Methodology (PM2) is the standard model currently used by the Army. The assumptions for PM2 are found in Ellner and Hall,[1] and the resulting failure rate after some amount of testing, t, can be expressed as

$$\lambda_G = (1 - MS)\lambda_I + \frac{(1-\mu_d)MS\lambda_I\beta t}{(1+\beta t)} + \frac{MS\lambda_I}{(1+\beta t)} . \tag{8}$$

By setting t in Eq. 8 to the amount of reliability growth test time, T, the planning curve can be fully determined. The final parameter, β, can be expressed in terms of the previously defined parameters and is given in Ellner and Hall[1] as

$$\beta = \left(\frac{1}{T}\right)\left(\frac{1-\frac{M_I}{M_G}}{\mu_d MS-\left(1-\frac{M_I}{M_G}\right)}\right). \tag{9}$$

3. METHODOLOGY

The test program is assumed to consist of three phases: reliability growth, reliability demonstration, and reliability assurance testing. Reliability demonstration and assurance testing are both fixed configuration tests that are used to provide an overall reliability assessment. The demonstration test results are used as prior information and combined with the results from the assurance test. The main difference between them is that the assurance test is a system-level operational test, while the demonstration test may use non-military operators and can also be performed at either the system or subsystem level. Reliability growth testing is then performed to support the demonstration test. The intent of the growth testing is to ensure that the system has a high probability of achieving sufficient reliability prior to demonstration testing. Sufficient reliability is defined as seeing at most the allowable number of failures in demonstration and assurance testing, thereby successfully meeting the reliability requirement for the system.

The end goal of the reliability program is to enter the RDT and follow-on assurance test with sufficient reliability to successfully pass the assurance test. This means that there should be a high probability of observing no more than the predetermined desired number of failures during the RDT and assurance tests. The reliability growth phase should therefore be structured to support achieving the desired results during RDT and the subsequent assurance test.

3.1 Assurance Testing

3.1.1 Definition of Posterior Risks

The posterior consumer's risk is defined as the probability of not meeting the requirement, conditioned on the test being passed.[2] This is given as

$$p(\lambda > \lambda_R \mid n_s \leq c) = \int_{\lambda_R}^{\infty} \frac{p(n_s \leq c|\lambda)p(\lambda)}{\int_0^{\infty} p(n_s \leq c|\lambda)p(\lambda)d\lambda} d\lambda, \tag{10}$$

where n_s is the observed number of failures, c is the allowable number of failures, $p(\lambda)$ is the prior distribution for λ, and λ_R is the failure rate requirement. Note that this definition of consumer risk does not directly result in a specified posterior probability relative to the failure rate requirement unless c = 0. Substituting $n_s = c$ in the right-hand side of Eq. 10 yields the actual posterior probability of observing c failures. The resulting posterior percentile can serve as an alternative consumer risk metric when a specific posterior probability is desired, and it should in most cases provide a more conservative result.

The posterior producer's risk is defined as the probability of having a failure rate below the requirement, conditioned on the test being failed. It is defined as

$$p(\lambda \leq \lambda_R \mid n_s > c) = \int_0^{\lambda_R} \frac{p(n_s > c|\lambda)p(\lambda)}{\int_0^\infty p(n_s > c|\lambda)p(\lambda)d\lambda} d\lambda \, , \tag{11}$$

with the same terms as in the posterior consumer's risk. Note that the requirement, λ_R, can also be replaced by λ_R/d, where d is defined as the discrimination ratio. Applying the discrimination ratio allows for the producer risk to be calculated using a failure rate value somewhat less than the requirement itself, which often presents more flexibility when choosing an appropriate assurance test design.

An additional useful measure is the overall probability of passing the test, which is given as

$$p(n \leq c) = \int_0^\infty p(n \leq c|\lambda)p(\lambda)d\lambda \, . \tag{12}$$

It is also important to note that the prior defined by $p(\lambda)$ is implicitly defined as the posterior $p(\lambda|x)$ resulting from modeling using previous data generically defined by x. Although the expressions for the posterior risks can be calculated directly for the simple prior-likelihood combination presented here, it is also straightforward to use posterior samples from simulation techniques such as Markov chain Monte Carlo (MCMC) to calculate the risks. Each posterior sample can be used in the conditional terms in Eqs. 10, 11, and 12, which allows for risk calculations in cases where the prior-likelihood combination is more complex. More information on the posterior risks, along with additional forms for other test and data assumptions, is also given in Hamada et al.[2]

3.1.2 Determining Assurance Test Lengths

For many defense applications, the main objective of the assurance test will be to ensure that the desired percentile of the posterior MTBF for the system is at or above the MTBF requirement after a successful test. Designing a test with a high probability of success is also important, as it would be impractical to design a test that met posterior risk thresholds while still having a low chance of being successfully passed. This is a particularly important point in programs that have severe constraints on resources. Controlling the posterior risks while simultaneously ensuring that the probability of passing the test remains high should result in a test plan that provides an efficient means of assessing reliability in these resource-constrained cases.

Using the available prior information, the process starts with a desired number of allowable failures and minimum desired test length. The test length (and if necessary, the allowable failures) are then varied until the posterior risks and probability of passing conditions are achieved. Note that this condition is similar, but not identical, to stating that the posterior consumer's risk must be less than or equal to 0.20. The following procedure can be used to determine the required test length and number of allowable failures for the assurance test[2]:

1. Generate samples from posterior distribution on failure rate for the assumed RDT data.
2. Generate the posterior predictive number of failures for each sampled value of the posterior failure rate.
3. Posterior consumer risk will be the proportion of sampled failure rates that are above λ_R when the corresponding number of failures is less than or equal to (or alternatively, just equal to) the current allowable number. Calculate the test length required for the desired posterior consumer risk.
4. Posterior producer risk will be the proportion of sampled failure rates that are below λ_R when the corresponding number of failures is greater than the current allowable number. If producer risk is greater than the desired threshold, increase the allowable number of failures and return to step 3.

Steps 3 and 4 can be repeated until the number of allowable failures and test length meet the desired thresholds for both posterior risks. The overall probability of passing the test defined in Eq. 12 can also be calculated using the posterior predictive samples generated in steps 1 and 2. The probability of passing is simply the proportion of sampled failures that are less than or equal to the allowable number of failures. This value should be considered along with the posterior risks, as it is possible to design assurance tests that meet desired risk thresholds while having little chance of passing.

3.2 Assurance Test Development Examples

This section presents two example assurance test plans using two different Bayesian models for combining the RDT and assurance test results. The first is a straightforward Gamma–Poisson model, and the second is a hierarchical extension of the Gamma–Poisson.

3.2.1 Gamma–Poisson Model

The MTBF requirement for the example is assumed to be 1000 h, and the RDT is assumed to be 4,000 h. The model for combining the RDT and assurance test results first assumes a constant rate of occurrence of failure defined by λ. The likelihood for *n* observed failures in RDT test length *T* is then assumed to be Poisson(λ), defined as

$$l(n|\lambda, T) = \frac{(\lambda T)^n}{n!} exp(-\lambda T). \tag{13}$$

For the prior distribution on λ, assume λ is Gamma(α, τ) such that

$$p(\lambda) = \frac{\lambda^{\alpha-1}\tau^{\alpha}}{\Gamma(\alpha)} exp(-\tau\lambda). \tag{14}$$

The prior and likelihood defined by Eqs. 13 and 14 yield a Gamma posterior given by

$$p(\lambda|n) = \frac{\lambda^{n+\alpha-1}(\tau+T)^{\alpha+n}}{\Gamma(\alpha+n)} exp[-(\tau+T)\lambda]. \tag{15}$$

Examining the distribution in Eq. 15 shows that small values for the α and τ parameters (i.e., 0.001 or smaller) will provide a weak prior distribution, allowing the posterior to be defined mainly by the observed failures and test length.

Defining n_s as the number of system failures observed in the assurance test of length T_s, samples from the distribution in Eq. 15 can then be used to calculate the posterior risks associated with the reliability assurance test. The posterior risks can also be expressed analytically for this case, allowing for more efficient calculation. The posterior consumer risk will be defined as

$$p(\lambda > \lambda_R \mid n_s \leq c) = \frac{\sum_{i=0}^{c} \frac{{T_s}^i(\tau+T)^{\alpha+n}\Gamma(\alpha+n+i,(\tau+T+T_s)\lambda_R)}{i!\,\Gamma(\alpha+n)(\tau+T+T_s)^{\alpha+n+i}}}{\sum_{i=0}^{c} \frac{{T_s}^i(\tau+T)^{\alpha+n}\Gamma(\alpha+n+i)}{i!\,\Gamma(\alpha+n)(\tau+T+T_s)^{\alpha+n+i}}}. \tag{16}$$

The posterior distribution for the exact outcome in the assurance test is also a Gamma distribution, where the upper posterior percentile for λ_R is given by

$$p(\lambda > \lambda_R \mid n_s) = \frac{\Gamma(\alpha+n+n_s,(\tau+T+T_s)\lambda_R)}{\Gamma(\alpha+n+n_s)}. \tag{17}$$

Though different from the formal definition for posterior consumer's risk in Hamada et al.,[2] the expression in Eq. 17 is an alternate form for the probability of not meeting the requirement, conditioned on the test being passed. This definition may be preferred in many settings, where the desire after the assurance test is to have a relatively small posterior probability that the system's failure rate is higher than the actual requirement. As demonstrated in Table 1 and the example application in Section 5, the use of Eq. 17

will guarantee this outcome, while the form in Eq. 16 may result in posterior probabilities that are higher than desired.

Table 1 shows a comparison of the results for the different Consumer Risk definitions discussed in Section 3.1.1. The RDT is held constant at 4,000 h with four failures, and the MTBF requirement is 1,000 h (i.e., λ_R= 0.001). Aside from the zero-failure test where the two risk definitions are equal, the test length and allowable failures using the standard Posterior Consumer Risk will result in a posterior probability that is well above common thresholds for consumer risk (e.g., 0.20). This highlights the difference between the standard Posterior Consumer Risk definition and the actual posterior outcome for the test.

Table 1. Assurance test results comparing standard posterior consumer risk definition with posterior probability

RDT Hours	RDT Failures	Assurance Test Hours	Assurance Test Failures	Posterior Consumer Risk	Posterior Probability $\lambda > \lambda_R$
4000	4	1515	0	0.20	0.20
4000	4	2272	1	0.20	0.25
4000	4	3055	2	0.20	0.29
4000	4	3834	3	0.20	0.33
4000	4	4632	4	0.20	0.37

It is desirable to have a small posterior probability of being above the requirement failure rate, so the consumer risk is defined for this example using the actual posterior distribution. This will ensure that the resulting posterior probability bound on the failure rate aligns with the chosen risk threshold. The desired thresholds are set to 0.20 for both consumer and producer risks, and the discrimination ratio is also set at 1.5. This discrimination ratio value implies that if the test is failed, there is a low probability of the system's MTBF (reciprocal of the failure rate) being more than 50% above the MTBF requirement.

The posterior producer risk for this model is given as

$$p(\lambda \leq \lambda_R \mid n_s > c) =$$

$$\frac{1 - \frac{\Gamma(\alpha+n, (\tau+T)\lambda_R)}{\Gamma(\alpha+n)} - \sum_{i=0}^{c} \frac{T_s{}^i (\tau+T)^{\alpha+n} \Gamma(\alpha+n+i)}{i!\,\Gamma(\alpha+n)(\tau+T+T_s)^{\alpha+n+i}} \left(1 - \frac{\Gamma(\alpha+n+i, (\tau+T+T_s)\lambda_R)}{\Gamma(\alpha+n+i)}\right)}{1 - \sum_{i=0}^{c} \frac{T_s{}^i (\tau+T)^{\alpha+n} \Gamma(\alpha+n+i)}{i!\,\Gamma(\alpha+n)(\tau+T+T_s)^{\alpha+n+i}}}. \quad (18)$$

It is useful to examine a range of potential failure outcomes for the demonstration test and the corresponding impacts on the reliability assurance test. Table 2 contains the resulting assurance test lengths and allowable failures for five cases, which correspond to RDT outcomes ranging from one to four failures.

Table 2 Example demonstration and assurance test results for Gamma‑Poisson model

Case #	RDT Hours	RDT Failures	Assur. Test Hrs.	Assur. Test Failures	Posterior Prob. $\lambda > \lambda_R$	Posterior Producer Risk	Prob. of Passing
1	4000	1	3905	5	0.20	0.20	0.99
2	4000	2	3903	4	0.20	0.18	0.90
3	4000	3	2723	2	0.20	0.19	0.67
4	4000	4	2723	1	0.20	0.15	0.33

The results in Table 2 show a range of feasible assurance test plans, each of which meet the desired risk thresholds. The probability of passing varies significantly though, and it should be considered in the selection of an assurance test plan.

3.2.2 Hierarchical Gamma–Poisson Model

The hierarchical extension of the Gamma–Poisson model uses the same prior and likelihood given in Eqs. 13 and 14, along with additional hierarchical prior distributions on the parameters of the prior Gamma distribution.

The hyperparameters of the Gamma distribution are also assumed to follow a Gamma distribution. The hyperprior in Eq. 19 is assumed to be identical for both α and τ, with $\tilde{\alpha} = \tilde{\tau} = 0.001$. The small parameter values create a weak hyperprior, which reduces the impact on the overall resulting posterior distribution. Though not usually necessary, this restriction can also be relaxed via a straightforward extension if different prior distributions are desired.

$$p(x|\tilde{\alpha},\tilde{\tau}) = \frac{x^{\tilde{\alpha}-1}\tilde{\tau}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} exp(-\tilde{\tau}x) \tag{19}$$

The joint proportional posterior distribution for the individual failure rates, λ_i, and the Gamma parameters, α and τ, is then given as

$$p(\vec{\lambda},\alpha,\tau|\,\vec{n},\vec{T}) \propto \alpha^{\tilde{\alpha}-1}exp(-\tilde{\tau}\alpha)\tau^{\tilde{\alpha}-1}exp(-\tilde{\tau}\tau)\prod_{i=1}^{m}\frac{\lambda_i^{\alpha-1}\tau^{\alpha}}{\Gamma(\alpha)}exp(-\tau\lambda_i){\lambda_i}^{n_i}exp(-\lambda_i T_i)\,, \tag{20}$$

where $\vec{n} = (n_1,\dots,n_m)$ and $\vec{T} = (T_1,\dots,T_m)$ are vectors containing the number of failures and total test times for each of the *m* test events.

The posterior in Eq. 20 can be solved using MCMC. A Slice-in-Gibb's sampler can be easily applied, where each conditional distribution in the Gibb's sampler is estimated using a slice sampling technique.[3] The main benefit of the slice sampler is that it provides a straightforward method for sampling from the desired distributions without the need for the tuning involved in traditional Metropolis-in-Gibb's approaches. Each of the distributions can also be shown to have a single mode, which greatly simplifies the slice sampling routine when estimating the full posterior. The conditional distributions for the algorithm are given in Eqs. 21–23. The distributions in Eqs. 21 and 23 are Gamma distributions, which greatly simplify the Gibb's sampler.

$$p\left(\lambda_i | \lambda_{j\neq i}, \alpha, \tau, \vec{n}, \vec{T}\right) \propto \lambda_i^{\alpha+n_i-1} exp[-(\tau+T_i)\lambda_i] \tag{21}$$

$$p\left(\alpha | \vec{\lambda}, \tau, \vec{n}, \vec{T}\right) \propto \alpha^{\tilde{\alpha}-1} exp(-\tilde{\tau}\alpha) \prod_{i=1}^{m} \frac{\lambda_i^{\alpha-1}\tau^{\alpha}}{\Gamma(\alpha)} \tag{22}$$

$$p\left(\tau | \vec{\lambda}, \alpha, \vec{n}, \vec{T}\right) \propto \tau^{\tilde{\alpha}+m\alpha-1} exp\left[-\left(\tilde{\tau} + \sum_{i=1}^{m} \lambda_i\right)\tau\right] \tag{23}$$

Table 3 shows the analogous results to those found in Table 2 for the hierarchical model. The same risk thresholds are used, and the discrimination ratio is again set to 1.5. The hierarchical model yields shorter assurance test lengths, but the reduction comes at the expense of significantly lower probabilities of passing the test. Use of the hierarchical model in this setting should be carefully examined, and it may prove to be more desirable in scenarios where there are multiple phases of prior test data available.

Table 3. Example demonstration and assurance test results for hierarchical Gamma–Poisson model

Case #	RDT Hours	RDT Failures	Assur. Test Hrs.	Assur. Test Failures	Posterior Prob. $\lambda > \lambda_R$	Posterior Producer Risk	Prob. of Passing
1	4000	1	1839	1	0.20	0.02	0.50
2	4000	2	1911	1	0.20	0.02	0.49
3	4000	3	1994	1	0.20	0.02	0.47
4	4000	4	2032	1	0.20	0.02	0.46

4. RELIABILITY GROWTH PLANNING AHEAD OF THE ASSURANCE TEST

The reliability demonstration and assurance testing can be aligned with a reliability growth plan by choosing a goal reliability that provides a suitably high probability of achieving the desired results for the demonstration and assurance testing. The approach can be applied at the system or subsystem level, depending on the desired testing and resources available during the reliability growth program. The first step in the process is to define the probability of seeing the desired results in the demonstration test (i.e., the number of observed failures is less than or equal to the allowable failures). From Wayne and Modarres,[4] the goal failure rate will follow a Gamma(α, τ) distribution. For an RDT of length T, the probability of seeing at most n allowable failures is given as

$$p(x \leq n) = \sum_{i=0}^{n} \frac{T^i \tau^\alpha \Gamma(\alpha + i)}{i!\,\Gamma(\alpha)(\tau + T)^{(\alpha+i)}} \,. \tag{24}$$

The α and τ are both unknown Gamma parameters in Eq. 24, but the τ can be expressed completely in terms of common reliability growth planning parameters. Choosing a desired value for the probability on the left-hand side of Eq. 24 then determines the value of α. This step fully defines the distribution of the goal failure rate and completes the connection between the reliability growth plan and the combined demonstration and assurance testing.

The expected value and variance of the goal failure rate at time t, $\lambda_G(t)$, are expressed using reliability growth parameters[5] and shown in Eqs. 25 and 26.

$$E[\lambda_G(t)] = \frac{\lambda_A \beta t}{1 + \beta t} + \frac{(1 - \mu_d)\lambda_B \beta t}{1 + \beta t} + \frac{\lambda_I}{1 + \beta t} \tag{25}$$

$$Var[\lambda_G(t)] = \frac{\lambda_A t}{\left(\frac{1}{\beta} + t\right)^2} + \frac{(1 - \mu_d)^2 \lambda_B t}{\left(\frac{1}{\beta} + t\right)^2} + \frac{\lambda_I}{\beta\left(\frac{1}{\beta} + t\right)^2} \tag{26}$$

For the expressions in Eqs. 25 and 26, λ_I is the total initial failure rate for the system at the start of testing, λ_A is the failure rate for failure modes that will not have corrective actions applied after discovery, λ_B is the failure rate for failure modes that will have corrective actions, μ_d is the average FEF for the corrective actions, and β is an additional parameter associated with the underlying distribution of the failure rates for the failure modes in the system.[4]

Defining the expected value of the goal failure rate in Eq. 25 as λ_G and solving for the β parameter yield

$$\beta = \frac{\lambda_I - \lambda_G}{t[\lambda_G - \lambda_I(1-\mu_d MS)]} \, , \tag{27}$$

which is identical to the form provided earlier in Eq. 9.

Next, define the growth potential failure rate, λ_{GP}, which is the theoretical lower bound on the failure rate that can be achieved after all modes are observed and corrected with an assumed level of fix effectiveness. The growth potential failure rate can be expressed as

$$\lambda_{GP} = \lambda_I(1-\mu_d MS). \tag{28}$$

The reliability growth potential is often used to manage risk when planning a test program. It is not generally achievable in the limited testing associated with a reliability growth program, but the growth potential failure rate needs to be lower than the desired goal failure rate for the growth program. Combining Eq. 28 with the ratio of the goal MTBF to the growth potential MTBF defined in Eq. 5 then yields

$$\lambda_G = \frac{\lambda_I(1-\mu_d MS)}{R} \, . \tag{29}$$

Inserting the expression in Eq. 29 into Eq. 27 then provides an alternate expression for β in Eq. 30. β can be expressed in terms of management metrics, the desired ratio of the MTBF goal to the MTBF growth potential, and the reliability growth test time.

$$\beta = \frac{R - 1 + \mu_d MS}{t(1-R)(1-\mu_d MS)} \tag{30}$$

The definition of the expected value of the Gamma distribution can then be used to define the expected value of the goal failure rate in Eq. 29 in terms of the Gamma parameters. This yields the expression for α given in Eq. 31.

$$\alpha = \frac{\tau\lambda_I(1-\mu_d MS)}{R} \tag{31}$$

The variance of the goal failure rate can also be expressed using the Gamma parameters, and the variance definition along with the expression in Eq. 31 yields

$$Var[\lambda_G(t)] = \frac{\alpha}{\tau^2} = \frac{\lambda_I(1-\mu_d MS)}{\tau R}. \quad (32)$$

An additional expression for the variance of the goal failure rate can be found by inserting Eq. 30 into Eq. 26, which yields

$$\begin{aligned} Var[\lambda_G(t)] = \left(\frac{\lambda_I}{t}\right)\Bigg(&\left[\frac{R-1+\mu_d MS}{(1-R)(1-\mu_d MS)+R-1+\mu_d MS}\right]^2 MS[(1-\mu_d)^2-1] \\ &+\left[\frac{R-1+\mu_d MS}{(1-R)(1-\mu_d MS)+R-1+\mu_d MS}\right]\Bigg). \end{aligned} \quad (33)$$

Equating the right-hand sides in Eqs. 32 and 33 and solving for τ results in an expression in terms of management metrics, the desired ratio of the MTBF goal to the MTBF growth potential, and the reliability growth test time t.

$$\tau = \frac{t(1-\mu_d MS)}{R\left(\left[\frac{R-1+\mu_d MS}{(1-R)(1-\mu_d MS)+R-1+\mu_d MS}\right]^2 MS[(1-\mu_d)^2-1]+\left[\frac{R-1+\mu_d MS}{(1-R)(1-\mu_d MS)+R-1+\mu_d MS}\right]\right)} \quad (34)$$

The expression for τ can then be used directly in the desired probability defined in Eq. 24, which can then be solved numerically for the necessary value of α. With both α and τ known, the expected value of the goal failure rate is known. The definition in Eq. 25 can also be used to determine the initial failure rate for the start of the reliability growth planning curve, which is given as

$$\lambda_I = \frac{\alpha(1+\beta t)}{\tau[\beta t(1-\mu_d MS)+1]} \quad (35)$$

This value, along with the management metrics and the previously defined β, uniquely determines the reliability growth planning curve. The goal failure rate for the curve will also have a distribution that yields the desired probability of seeing no more than the predefined number of failures for the RDT. Note that the expected value of the failure rate developed using the process defined here will provide a traditional idealized planning curve as defined in Wayne and Modarres.[4] Additional steps can be added to the curve for individual test phases during reliability growth testing. The expected values of other useful metrics (e.g., number of failure modes observed, percent of initial failure rate observed) can also be calculated using the formulas for traditional planning curves defined in Wayne and Modarres.[4]

5. EXAMPLE APPLICATION

The scenario for this example involves testing throughout the reliability growth, demonstration, and assurance events. The demonstration and assurance tests are combined using the simple Gamma–Poisson model from Section 3.2.1. The reliability growth testing is assumed to be 16,550 h. Table 4 contains the resulting initial MTBF, goal MTBF, and the expected number of failure modes for each of the cases in Table 2. The probability of seeing the RDT failures defined in Eq. 24 is set to 0.7 for each case. The MS and FEF are also assumed to be 0.95 and 0.75, respectively, and the ratio of the goal MTBF to the growth potential MTBF is also set to 0.7.

Table 4. Reliability growth values for RDT and assurance test results

Case #	RDT Hours	RDT Failures	Assurance Test Hours	Assurance Test Failures	Goal MTBF	Initial MTBF	Expected Number of Failure Modes
1	4000	1	3905	5	3621	1487	4.1
2	4000	2	3903	4	2095	860	7.1
3	4000	3	2723	2	1454	597	10.2
4	4000	4	2723	1	1107	455	13.4

Figure 2 shows the resulting reliability growth planning curve for Case 2 from Table 4. The reliability growth testing is divided into three equal phases labeled as RGT1, RGT2, and RGT3, with corrective action periods assumed between each of them. The Idealized Curve represented by the black line represents the reliability growth that would occur if testing was paused upon discovery of a failure mode and not resumed until a corrective action was implemented for the observed failure mode. The corrective action periods assumed between the phases instead imply that the corrective actions are implemented in bulk at predetermined points during the reliability growth testing. Implementation of corrective actions during these periods allows the MTBF to grow from the initial value of 860 h to the goal value of 2,095 h in three steps, which reduces the risk associated with delaying all corrective actions until prior to the RDT. As shown in Table 4, the plan implies that approximately seven failure modes are expected to be observed during the reliability growth testing. Other metrics associated with the reliability growth plan are also available, and they are identical to those defined by the PM2 Reliability Growth Planning Model.[1]

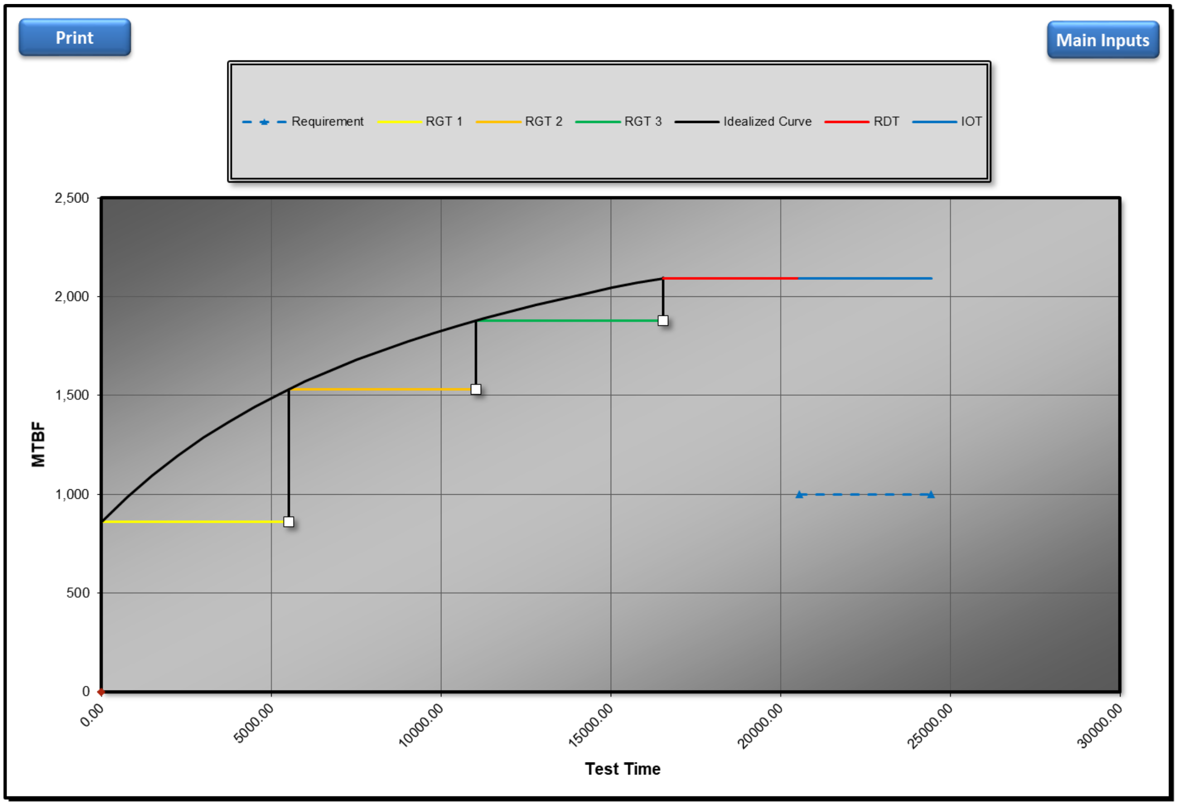

Figure 2. Sample reliability growth planning curve using assurance testing

For comparison purposes, Figure 3 shows the analogous result using traditional demonstration testing with OC curves in the PM2 model. The OC curve analysis assumes the consumer risk is 0.20 and the probability of acceptance is 0.90, which is analogous to the values used in the assurance testing example. The resulting IOT test length is 6,721 h for the same four-failure test. This value is significantly higher than the analogous assurance test length of 3,903 h from Table 2. The curve also requires significantly higher MTBF values to achieve the statistical risks assumed for the OC curve analysis. The MTBF goal increases from 2,095 to 2,763 h, and the initial MTBF increases from 860 to 1,135 h. These results clearly show the potential benefits associated with the assurance testing approach, which reduces the technical risk of achieving higher MTBF targets while also increasing the overall efficiency of the test program.

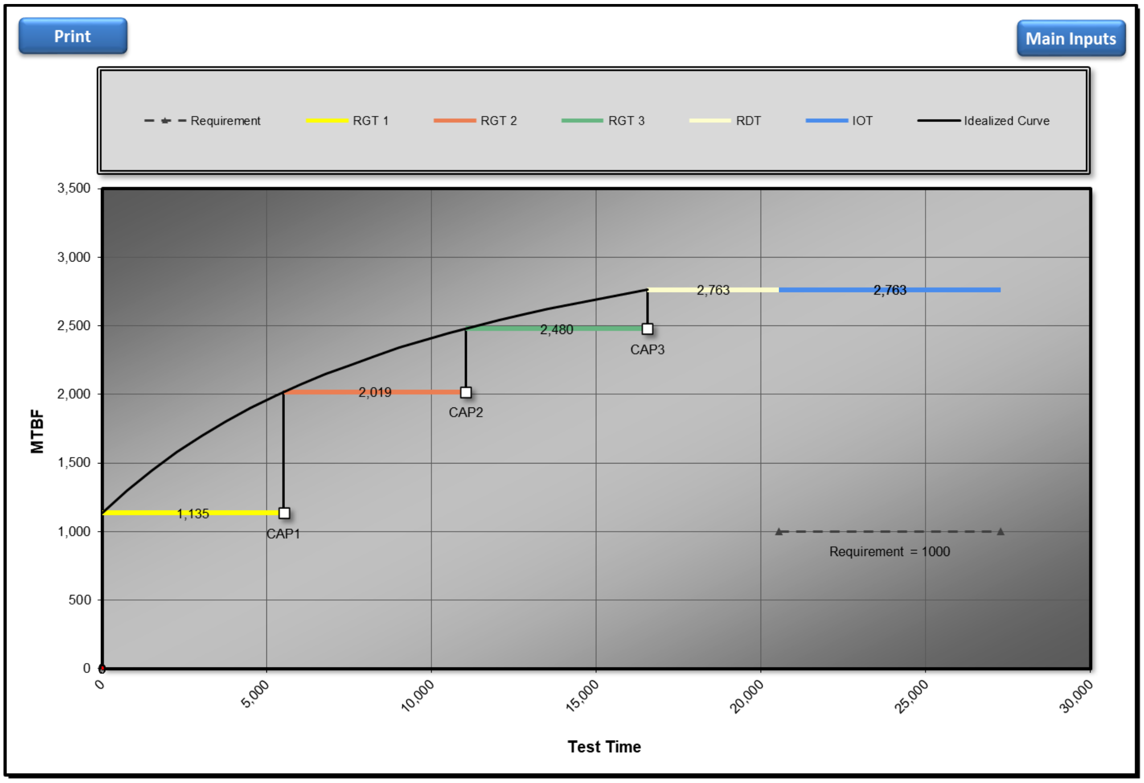

Figure 3. Sample reliability growth planning curve using demonstration testing with OC curves

6. CONCLUSIONS AND FUTURE WORK

Extensions to the work presented here are possible. The most obvious is the application of the assurance testing results to subsystem testing, where the prior information from the RDT is obtained for subsystems and then combined with system-level results in the assurance test. An additional area for further examination is the application of prior distributions to the management metrics (i.e., MS and FEF) to allow for a fuller treatment of the uncertainty that may exist in the reliability growth plan. Both efforts will provide a more complete treatment of the subject and allow for broader application across Army and other defense systems.

7. REFERENCES

1. Ellner, P.M., Hall, J.B. (2006). *An approach to reliability growth planning based on failure mode discovery and correction using AMSAA projection methodology.* Proceedings of the Annual Reliability and Maintainability Symposium.

2. Hamada, M., Wilson, A., Reese, C.S., & Martz, H. (2010). *Bayesian reliability*. Springer.

3. Neal, R.M. (2003). Slice sampling. *Annals of Statistics*, 31–3.

4. Wayne, M., & Modarres, M. (2015). A Bayesian model for complex system reliability growth under arbitrary corrective actions. *IEEE Transactions on Reliability*, 64–1.

5. Wayne, M. (2018). *Modeling uncertainty in reliability growth planning*. Proceedings of the Annual Reliability and Maintainability Symposium.

LIST OF ACRONYMS

AMSAA	U.S. Army Materiel Systems Analysis Activity
DAC	DEVCOM Analysis Center
DEVCOM	U.S. Army Combat Capabilities Development Command
FEF	Fix Effectiveness Factor
MCMC	Markov chain Monte Carlo
MS	Management Strategy
MTBF	Mean Time Between Failure
OC	Operating Characteristic
PM2	Planning Model Based on Projection Methodology
RDT	Reliability Demonstration Test

ORGANIZATION

DEVCOM Analysis Center
FCDD-DAS-L/M. Wayne
6896 Mauchly St.
Aberdeen Proving Ground, MD 21005-5071

DEVCOM Army Research Laboratory
FCDD-RLB-CI/Tech Library
2800 Powder Mill Rd.
Adelphi, MD 20783

Defense Technical Information Center
ATTN: DTIC-O
8725 John J. Kingman Rd.
Fort Belvoir, VA 22060-6218