

International Journal of
**RELIABILITY, QUALITY
AND SAFETY
ENGINEERING**

Volume 10 • Number 1 • March 2003

**Probability Models for
Sequential-Stage System Reliability
Growth via Failure Mode Removal**

D. P. Gaver, P. A. Jacobs, K. D. Glazebrook and E. A. S

 **World Scientific**
New Jersey • London • Singapore • Hong Kong • Bangalore

PROBABILITY MODELS FOR SEQUENTIAL-STAGE SYSTEM RELIABILITY GROWTH VIA FAILURE MODE REMOVAL

DONALD P. GAVER* and PATRICIA A. JACOBS†

Operations Research Department, Naval Postgraduate School, Monterey, California

* *dgaver@nps.navy.mil*

† *pjacobs@nps.navy.mil*

KEVIN D. GLAZEBROOK

Department of Business Studies, University of Edinburgh,

Edinburgh EH8 9JY, United Kingdom

kevin.glazebrook@ed.ac.uk

ERNEST A. SEGLIE

The Pentagon Washington, DC

eseglie@dote.osd.mil

Received 19 March 2002

Revised 11 December 2002

This paper provides guidance for the planners of a test of any system that operates in sequential stages: only if the first stage functions properly (e.g., a vehicle's starter motor rotates adequately) can the second stage be activated (ignition system performs) and hence tested, followed by a third stage (engine starts and propels vehicle), with further stages such as wheels, and steering, and finally brakes eventually brought to test. Each sequential stage may fail to operate because its design, manufacture, or usage has faults or defects that may give rise to failure. Testing of all stages in the entire system in appropriate environments allows failures at the various stages to reveal defects, which are targets for removal. Early stages' fault activations thus postpone exposure of later stages to test. It is clear that only by allowing the entire system to be tested end-to-end, through all stages, and to observe several total system successes can one be assured that the integrated system is relatively free of defects and is likely to perform well if fielded.

The methodology of the paper permits a test planner to hypothesize the numbers of (design) faults present in each stage, and the stagewise probability of a fault activation, leading to a system failure at that stage, *given* survival to that stage. If the test item fails at some stage, then rectification ("fix") of the design occurs, and the fault is (likely) removed. Failure at that stage is hence less likely on future tests, allowing later stages to be activated, tested, and fixed. So reliability grows.

To allow many Test and Fix (TAF) cycles is obviously impractical. A stopping criterion proposed by E. A. Seglie that suggests test stopping as soon as an uninterrupted run/sequence of r (e.g., 5) consecutive system successes has been achieved is studied quantitatively here. It is shown how to calculate the probability of eventual field success if the design is frozen and the system fielded after such a sequential stopping criterion is

achieved. The mean test length is also calculated. Many other calculations are possible, based on formulas presented.

Keywords: System reliability; reliability growth; optimal stopping rule; probability models; design burn-in.

1. Introduction

This paper provides mathematical models of *reliability growth by design defect* or *failure-mode identification and removal* in system reliability testing and management, for instance during military Test and Evaluation (see Ref. 31). This is sometimes referred to as Test-and-Fix (TAF). The models demonstrate how testing can promote early learning about, and rectification of, system defects in design, manufacturing, and operations. In the military and elsewhere, such testing should, and does, begin with engineering-level Developmental Testing (DT) – initially of subsystems – and terminates with end-to-end Operational Testing (OT). At present, attempts are being made to compress and combine DT and OT so as to shorten acquisition time and decrease its expense. The models proposed are intended to provide insight to modern test planners. Software that exercises the models is available from the authors.

The model structure to be studied is the following: a system, \mathbf{S} , is made up of S ($S \geq 1$) subsystems or stages, each of which must function on demand, *in sequence*, for perfect operation; failure of any subsystem, *especially to interact with another subsystem* (interaction can also be viewed as a stage), means total system failure. Demands for subsystem, or inter-subsystem, function occur in order, stagewise; $s = 1, 2, \dots, S$. If a demand at an intermediate stage/subsystem, s , succeeds, i.e., any defects do not activate, a demand is placed on stage/subsystem $s + 1$; if all such demands succeed, the entire system operates successfully on that particular test or usage occasion (it may not again if remaining defects activate). Such sequentially activated systems are encountered, *inter alia*, in the testing of complex weapons and in software reliability.

To perform a system-level operational test of \mathbf{S} , suitable test conditions are first established. It is desirable to quantify those conditions (weather and other environmental effects, pre-test transport and setup stress, target properties, etc.). This can be done by incorporating explanatory variables to represent between-test variations. For recent related modeling, see Ref. 9. Under given conditions let each subsystem possess a certain (random, or at least unknown) number of failure modes (or defects), d_s , for subsystem s , $s \leq S$. These modes *become active* (cause failure) with probability θ_s if a demand is received at that stage; otherwise are inactive or survive with probability $1 - \theta_s = \bar{\theta}_s$. In order for the s th subsystem to *experience test*, and hence possibly reveal a failure mode, all previous $i \in (1, 2, \dots, s - 1)$ subsystems, and their interconnection and transition actions, must survive, and hence transmit, demands. If a failure mode in a subsystem is *activated* (causes failure), the design or execution may well be

modified. Here it is optimistically assumed that the failure mode is removed, and thus “reliability growth” occurs. This simplicity may not hold: new failure modes may actually be introduced, and bedrock non-removable failure modes will remain. Extensions of the current work to incorporate such features are in progress.

The ideas we explore are related to, but not the same as classical *burn-in*; whereby early testing removes weak components from an existing population; see Refs. 8 and 25. Our problem emphasizes *design burn-in*: systems are tested and the design improved *before* a population of manufactured and fielded items is created. Members of that population can possibly then experience classical burn-in before fielding, but the need should be reduced if the *design* has already been improved.

The paper is structured as follows: Sec. 2 presents some issues which are operationally relevant and which provide the focus for much of the later material. Section 3 introduces our models and develops procedures for the computation of key operating characteristics of given test designs. In Sec. 4 we propose a simple approach to the modeling of between test variability. A sequential test plan that ensures that all the stages will be tested at least r times is to test until there is a *run of r consecutive system successes*; (see Ref. 28). Section 5 discusses a Bayesian model that suggests that, while not being Bayes optimal in a formal sense, a runs test provides a simple and effective test stopping rule. All of the above discussions are complemented by numerical examples. The paper concludes with a summary in Sec. 6.

2. Operationally Relevant Questions

Given preliminary values of the parameters, inferred from engineering design and experience with analogous subsystems and systems, it is operationally meaningful to address such questions as these prior to expensive field testing:

- (a) After a given number of *system* tests, what is the (approximate) probability that the system will operate satisfactorily when released to the field or delivered to a user?
- (b) How many tests are likely to be required to achieve the first (or j th) end-to-end success?
- (c) How many tests are required to achieve r *consecutive* end-to-end test successes, or, in statistical parlance, a (first) *run of r* ? This is a possible test-stopping rule that is attractive because of its simplicity and intuitive evidence of system success.
- (d) Suppose testing is stopped after T tests, where T is determined by some stopping rule, such as in (c). After testing, no further design modifications are contemplated. What are the failure characteristics of the system if fielded: e.g., what is the *operational/field* probability of system (reliability) success? For a previous account of this measure of system success under “reliability growth”, see Ref. 15.

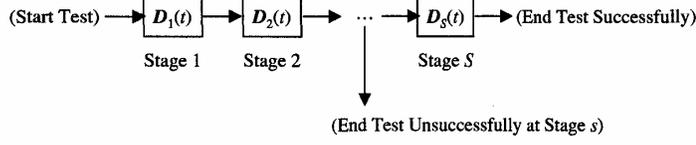


Fig. 1.

3. Models for Discovery of Hidden, or Sequentially-Evident, Design Defects

A system is composed of a number, s , of subsystems or stages, each of which contains an uncertain number of failure modes (design defects). We write $\mathbf{D}_s(t)$ for the number of defects present in stage s after t tests. When a design defect is activated during a test, the system fails at the corresponding stage, the defect is removed and that particular test terminates. Figure 1 illustrates the configuration and outcomes.

If test $t + 1$ is ended unsuccessfully at Stage s then

$$\mathbf{D}_{s'}(t + 1) = \begin{cases} \mathbf{D}_{s'}(t), & s' \neq s \\ \mathbf{D}_{s'}(t) - 1, & s' = s \end{cases}$$

while if test $t + 1$ is successful then $\mathbf{D}(t + 1) = \mathbf{D}(t)$. Note that the discussion in this section and the next, conditions upon the value of $\mathbf{D}(0)$, the configuration of design defects present before testing. In Sec. 5 we sketch a Bayesian approach to uncertainty about $\mathbf{D}(0)$.

We shall assume that, when subject to test, successive stages fail (or not) independently of each other. Conditional upon stage i having d_i defects the probability that a defect is revealed during test is $\tilde{p}_i(d_i) = 1 - \tilde{q}_i(d_i)$, where we take $\tilde{q}_i(0) = 1$. An important special case is the binomial model $\tilde{q}_i(d_i) = \theta_i^{d_i}$ for some $\theta_i \in (0, 1)$. However, allowance for extra variability in θ_i can be made by taking $\tilde{q}_i(d_i) = E[\theta_i^{d_i}]$ where $\theta \sim G_i$ for a distribution function G_i . This is natural to reflect within-stage variability beyond the simple binomial and may be viewed as representing physical mixtures. Although we assume that discovered defects are removed with certainty, we can incorporate a probability ρ_i of successful defect removal at stage i upon replacing $(1 - \theta_i)$ by $(1 - \theta_i)\rho_i$ and proceeding. We also use $\tilde{Q}_i(d_i)$ for the probability that no defect in stage i is revealed when the system is put in use *in the field* during a single mission.

3.1. Examples of conditional models $\tilde{q}(d)$

Here we drop the stage suffix i and explore possible choices for the over-variable model $\tilde{q}(d) = E[\theta^d]$ where $\theta \sim G$. If we take $\theta \sim \beta(a, b)$ then

$$\tilde{q}(d) = \frac{\Gamma(a + b)\Gamma(b + d)}{\Gamma(b)\Gamma(a + b + d)}$$

with $\tilde{q}(d) = (d + 1)^{-1}$ for the special case of uniform on $(0, 1)$. One possible normalization is to put the mean of the beta distribution when $d = 1$ equal to the original “deterministic” survival probability θ . Alternatively, using a gamma distribution with mean μ and shape parameter β upon $-\ln \theta$ yields

$$\tilde{q}(d) = \left(1 + \frac{\mu d}{\beta}\right)^{-\beta}$$

which coincides with the uniformly distributed θ case above when $\mu = \beta = 1$. Normalizing at $d = 1$ gives

$$\tilde{q}(1) = \left(1 + \frac{\mu}{\beta}\right)^{-\beta} = \theta \Rightarrow \mu = \beta(\theta^{-1/\beta} - 1),$$

where β is an optional tuning parameter. If we take $-\ln \theta$ to have a positive stable law (see Ref. 14) we obtain, upon normalization at $d = 1$, that

$$\tilde{q}(d) = \theta^{d^\beta}$$

for some $0 < \beta < 1$. Finally, when $-\ln \theta$ has an inverse Gaussian distribution, we obtain

$$\tilde{q}(d) = \exp\left\{-\frac{1}{c}\{(1 + 2c\mu d)^{1/2} - 1\}\right\},$$

where μ and c are, respectively, the mean and coefficient of variation of $-\ln \theta$.

3.2. Expected probability of system field success after a fixed number, t , of tests

Let $X\{\underline{\mathbf{D}}(t)\}$ be some function of the number of defects present in each stage after t tests. We adopt a forwards approach to the (recursive) computation of $E[X\{\underline{\mathbf{D}}(t)\}]$ as follows:

$$\begin{aligned} E[X\{\underline{\mathbf{D}}(t+1)\}] &= E[E[X\{\underline{\mathbf{D}}(t+1)\}|\underline{\mathbf{D}}(t)]] \\ &= E\left[X\{\underline{\mathbf{D}}(t)\} \prod_{i=1}^S \tilde{q}_i(\mathbf{D}_i(t)) + \left\{\sum_{j=1}^S X\{\underline{\mathbf{D}}(t) - \underline{\mathbf{1}}^j\}\right\} \right. \\ &\quad \left. \times [1 - \tilde{q}_j(\mathbf{D}_j(t))] \prod_{i=1}^{j-1} \tilde{q}_i(\mathbf{D}_i(t))\right] \end{aligned} \quad (3.1)$$

where $\underline{\mathbf{1}}^j$ denotes an S -vector with j th component equal to one and zeros elsewhere. Now consider the important special case

$$X\{\underline{\mathbf{D}}\} = I\{\underline{\mathbf{D}} = \underline{\mathbf{d}}\}$$

where I is the indicator function, such that

$$E[X\{\underline{\mathbf{D}}(t)\}] = P\{\underline{\mathbf{D}}(t) = \underline{\mathbf{d}}\} \equiv p(\underline{\mathbf{d}}, t),$$

the joint probability of the number of defects in each stage after t tests. From Eq. (3.1) we obtain the recursion

$$p(\underline{d}, t+1) = p(\underline{d}, t) \prod_{i=1}^S \tilde{q}_i(d_i) + \sum_{j=1}^S p(\underline{d} + \underline{1}^j, t) \{1 - \tilde{q}_j(d_j + 1)\} \left\{ \prod_{i=1}^{j-1} \tilde{q}_i(d_i) \right\} \quad (3.2)$$

which is initialized with

$$p(\underline{d}, 0) = I\{\underline{\mathbf{D}}(0) = \underline{d}\}. \quad (3.3)$$

The probability of system survival in the field after t tests is

$$\tilde{Q}(t) = E \left[\prod_{i=1}^S \tilde{Q}_i(\mathbf{D}_i(t)) \right] = \sum_{\underline{d}} p(\underline{d}, t) \left\{ \prod_{i=1}^S \tilde{Q}_i(d_i) \right\}. \quad (3.4)$$

The following result expresses a notion of reliability growth in this context.

Lemma 1. *If $\{\tilde{Q}_i(d), d \in N\}$ is a decreasing sequence and $\tilde{q}_i(d_i) < 1$ for all $d_i > 0, 1 \leq i \leq S$, then $\{\tilde{Q}(t), t \in N\}$ is an increasing sequence with $\lim_{t \rightarrow \infty} \tilde{Q}(t) = 1$.*

Proof. Let ω be any realization related to a sequence of $t+1$ tests, commencing from a situation with $\underline{\mathbf{D}}(0) = \underline{d}$. Use $\underline{\mathbf{D}}(t, \omega)$ and $\underline{\mathbf{D}}(t+1, \omega)$ to denote the configuration of defects remaining following t and $t+1$ tests respectively. Under the condition in the statement of the lemma, since $\underline{\mathbf{D}}(t, \omega) \geq \underline{\mathbf{D}}(t+1, \omega)$ component wise, then

$$\prod_{i=1}^S \tilde{Q}_i\{\mathbf{D}_i(t, \omega)\} \leq \prod_{i=1}^S \tilde{Q}_i\{\mathbf{D}_i(t+1, \omega)\} \quad (3.5)$$

Taking expectations we infer that $\tilde{Q}(t) \leq \tilde{Q}(t+1)$ and hence that $\{\tilde{Q}(t), t \in N\}$ is an increasing sequence. Since it is bounded above by 1, it must converge to a finite limit. That the latter must be 1 follows by use of the recursion (3.2) together with an induction argument. \square

The following numerical examples are based on four-stage systems with simple binomial stagewise failures: $\tilde{q}_i(d_i) = \tilde{Q}_i(d_i) = \theta_i^{d_i}$. We suppose that in all cases there are three defects present in each stage before testing. The quantity $\tilde{Q}(t)$ in Eq. (3.4) is plotted for three θ -configurations in Fig. 2.

The graph suggests that reliability growth in a several-stage serial system is not likely to have the characteristic of classical one-stage reliability growth models of Duane and later authors, e.g., Ref. 16. There are ample physical reasons for this behavior. They also imply that more rapid and complete defect elimination, and hence ‘‘reliability growth’’ occurs if the last-reached system stages are apt to fail sooner than the earlier-reached stages. The reason is that the need to re-test the last stages forces more tests of the earlier ones because of the end-to-end success requirement. It is unlikely that a designer or tester can ever directly influence such a

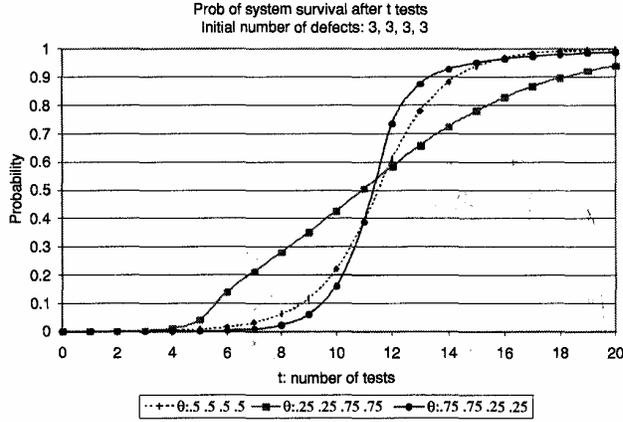


Fig. 2.

distribution of defects, but there may be implications for variations in the intensity of component-level testing: one might tolerate a few more defects in later stages, so that earlier stages will be subjected to more operational end-to-end tests.

3.3. Operating characteristics when testing stops after first run of r consecutive successful tests

Suppose the system test is stopped when there are $r \geq 1$ successful end-to-end tests in a row (a “first run of r ”). A test with this stopping rule ensures that all stages are tested at least r times. We use $\mathbf{T}(r)$ for the number of tests conducted until stopping and $\mathbf{D}\{\mathbf{T}(r)\}$ for the number of defects by stage, which are present then. Let $X[\mathbf{T}(r), \mathbf{D}\{\mathbf{T}(r)\}]$ be some function of these quantities. By conditioning upon the outcome of the first attempt to secure a run of r successes, and utilizing the Markovian structure of the process we obtain the backwards recursion

$$\begin{aligned}
 E[X[\mathbf{T}(r), \mathbf{D}\{\mathbf{T}(r)\}] | \mathbf{D}(0) = \underline{d}] &= X(r, \underline{d}) \left\{ \prod_{i=1}^S \tilde{q}_i(d_i) \right\}^r \\
 &+ \sum_{n=1}^r \sum_{j=1}^S E[X[\mathbf{T}(r) + n, \mathbf{D}\{\mathbf{T}(r)\}] | \mathbf{D}(0) = \underline{d} - \underline{1}^j] \\
 &\times \left\{ \prod_{i=1}^S \tilde{q}_i(d_i) \right\}^{n-1} \prod_{i=1}^{j-1} \tilde{q}_i(d_i) \{1 - \tilde{q}_j(d_j)\}.
 \end{aligned} \tag{3.6}$$

If we take

$$X[\mathbf{T}, \mathbf{D}] = \prod_{i=1}^S \tilde{Q}_i(\mathbf{D}_i), \tag{3.7}$$

then Eq. (3.6) provides a recursion for the quantity $p_r(\underline{d})$, the conditional expected probability of system mission survival in the field after the test, given that the initial number of defects is $\underline{\mathbf{D}}(0) = \underline{d}$. The recursion starts with $p_r(\underline{\mathbf{0}}) = 1$. Alternatively, if we take

$$X[\mathbf{T}, \underline{\mathbf{D}}] = \mathbf{T}, \quad (3.8)$$

then Eq. (3.6) provides a recursion for the quantity $\tau_r(\underline{d})$, the conditional expected number of tests until a run of r successes first occurs, given that the initial number of defects is $\underline{\mathbf{D}}(0) = \underline{d}$. An initial condition is $\tau_r(\underline{\mathbf{0}}) = r$.

The following result characterizes the behavior of the key operating characteristics $p_r(\underline{d})$ and $\tau_r(\underline{d})$.

Theorem 1. *If $\tilde{q}_i(d_i) < 1$ for all $d_i > 0$, $1 \leq i \leq S$, then it follows that for each fixed \underline{d}*

- (1) *The sequence $\{p_r(\underline{d}), r \in N\}$ is increasing with limit 1;*
- (2) *The sequence $\{\tau_r(\underline{d}) - r, r \in N\}$ is increasing with limit $\Delta(\underline{d})$ which satisfies the recursion*

$$\left\{ 1 - \prod_{i=1}^S \tilde{q}_i(d_i) \right\} \Delta(\underline{d}) = 1 + \sum_{j=1}^S \left(\prod_{i=1}^{j-1} \tilde{q}_i(d_i) \right) (1 - \tilde{q}_j(d_j)) \Delta(\underline{d} - \underline{\mathbf{1}}^j), \quad \underline{d} \neq \underline{\mathbf{0}},$$

$$\Delta(\underline{\mathbf{0}}) = 0.$$

Proof. We give the proof of (2) and omit the similar, but simpler, proof of (1). Let ω be any realization related to an infinite sequence of tests commencing from a situation in which $\underline{\mathbf{D}}(0) = \underline{d}$. Use $\mathbf{T}(r, \omega)$ for the number of tests required to obtain the first run of r successes. The quantity $\mathbf{T}(r, \omega) - r$ is then the number of tests prior to $\mathbf{T}(r, \omega)$ which do not contribute to the first run of r . Note that $\mathbf{T}(r, \omega) - r \leq \mathbf{T}(r+1, \omega) - (r+1)$ since any test which does not contribute to the first run of r cannot contribute to the first run of $r+1$. Taking expectations we conclude that for all r and \underline{d} $\tau_r(\underline{d}) - r \leq \tau_{r+1}(\underline{d}) - (r+1)$; hence the sequence $\{\tau_r(\underline{d}) - r; r \in N\}$ is increasing. It must, therefore, either have a finite limit or be divergent. To see that it cannot be the latter we note that, under the assumption of the theorem, for any current configuration of defects \underline{d} , the system has a geometrically distributed number of successes (with finite mean) before a defect is removed. Should all defects be removed at some point, a run of r successes is guaranteed. It follows that $\tau_r(\underline{d}) - r$ remains bounded and hence has a limit $\Delta(\underline{d})$, say. Direct use of Eq. (3.6) yields the recursion for $\Delta(\underline{d})$ in the statement of the theorem.

The recursion described in Eqs. (3.6) and (3.7) permits rapid numerical determination of the mean or unconditional probability of field success. However, simulations show that there can be considerable variability of the actual probability of field success around its expected value, depending on defect survival. The following forwards approach can be used to determine the full distribution of the probability of field survival after the test. Let $a_i \leq d_i$, $1 \leq i \leq S$, with \underline{d} the assumed initial

profile of numbers of defects. Write $\tilde{\gamma}_r(\underline{a})$ for the probability that at some point during testing the defect profile \underline{a} is encountered. Obtain the $\tilde{\gamma}_r(\underline{a})$ recursively from

$$\begin{aligned} \tilde{\gamma}_r(\underline{a}) = & \sum_{s=1}^S \tilde{\gamma}_r(\underline{a} + \underline{1}^s) \sum_{n=0}^{r-1} \left[\left\{ \prod_{\substack{i=1 \\ i \neq s}}^S \tilde{q}_i(a_i) \right\} \tilde{q}_s(a_s + 1) \right]^n \\ & \times \left\{ \prod_{i=1}^{s-1} \tilde{q}_i(a_i) \right\} [1 - \tilde{q}_s(a_s + 1)], \end{aligned} \quad (3.9)$$

with initial condition $\tilde{\gamma}_r(\underline{a}) = 1$. We then deduce that $\tilde{\gamma}_r(\underline{a})$, the probability of defect profile \underline{a} at the conclusion of the test, is given by

$$\gamma_r(\underline{a}) = \tilde{\gamma}_r(\underline{a}) \left\{ \prod_{i=1}^S \tilde{q}_i(a_i) \right\}^r. \quad (3.10)$$

The distribution of the probability of field success after the test may now be inferred from Eq. (10). □

3.3.1. Numerical results

In the example whose results are displayed in Fig. 3, we take $\tilde{q}_i(d_i) = \tilde{Q}_i(d_i) = \theta_i^{d_i}$. Note that Fig. 3 displays considerable robustness of mean mission survival outcome to number of design defects and activation probabilities: often the mission survival probability exceeds 0.8–0.9. Note that in the case $\theta: 0.75, 0.25, 0.75, 0.25$, the probability of mission survival after testing increases slightly as the initial number

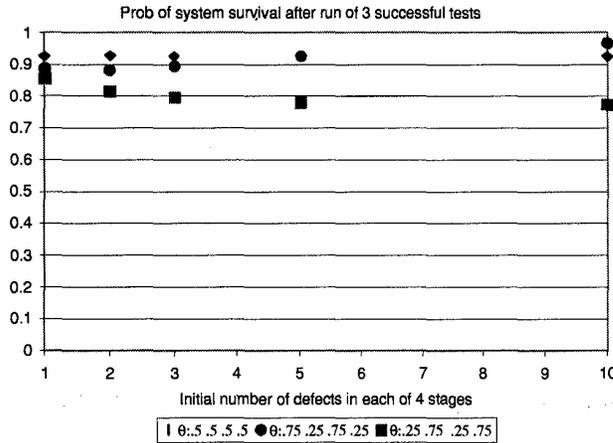


Fig. 3.

of defects in each stage increases. In this case the larger test activation probabilities $\bar{\theta}_i = 1 - \theta_i = 0.75$ for stages 2 and 4 result in more testing of stages 1 and 3. Consequently, the design defects in stages 1 and 3 are more apt to be discovered and removed; the probability of mission survival after testing increases as the number of defects increases.

In the examples presented in Figures 4–7 we compare key operating characteristics for the run's test for situations in which the conditional model for stage failure is (a) binomial with $\tilde{q}_i(d_i) = \theta_i^{d_i}$ and (b) an extra-variable model with $\tilde{q}_i(d_i) = E[\theta^{d_i}]$ where $\theta \sim \beta(a_i, b_i)$. We choose the parameters a_i, b_i , such that the mean of the beta distribution always equals θ_i in the corresponding comparator system.

From Figs. 4 and 6 it is striking that the *order* of the defect survival probability occurrence (which may be practically difficult to control at the developmental

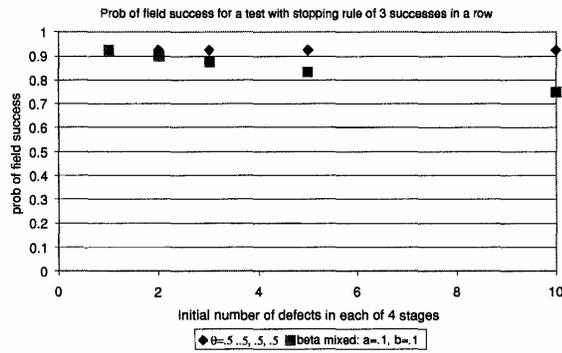


Fig. 4.

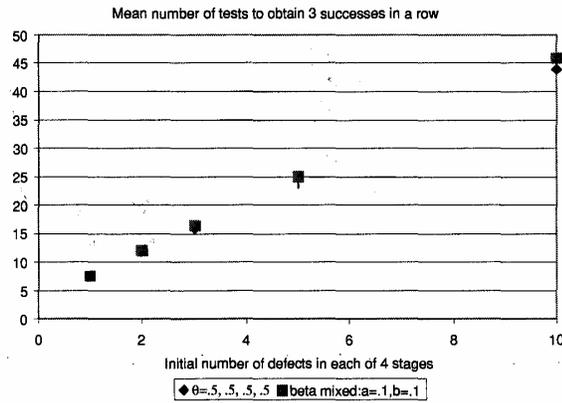


Fig. 5.

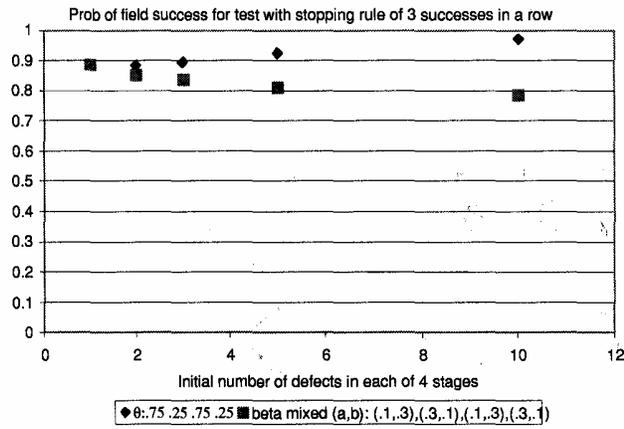


Fig. 6.

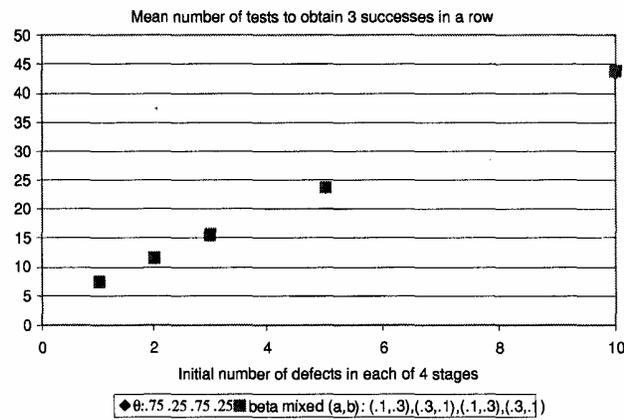


Fig. 7.

testing stage) can be influential at the final field survival probability level. Once again, in the case of Fig. 6, $\theta: 0.75, 0.25, 0.75, 0.25$ exhibits improved field response with *more* defects for the Bernoulli-trials case, but not for the over-variability case studied. Note that the mixing distribution employed in Fig. 4 is symmetric, but with high weighting near 0 and 1. Such an environment badly penalizes the tester if there are many (e.g., five or more) defects in the system initially. Figure 6 illustrates a similar phenomenon for another diffuse beta mixing distribution. From Figs. 5 and 7 it is seen that the *mean* times to achieve a success run of 3 for the different parametric cases are *remarkably similar*. These are isolated examples only, but certainly promote interest in run-like rules.

4. Examining the Effects of Test-to-Test Variability

We now suppose that we have an i.i.d. sequence $\varepsilon_1, \varepsilon_2, \dots$, of environmental test-specific random variables. These are assumed positive and held fixed for each entire test, with ε_n the variable applying to test n . The impact of these variables is as follows: if in test n , stage i has $\mathbf{D}_i(n)$ defects, then the probability that no defect is revealed should stage i be put to test is $[\tilde{q}_i\{\mathbf{D}_i(n)\}]^{\varepsilon_n}$. With this modification, Eq. (3.6), for example, is replaced by the recursion

$$\begin{aligned} E [X[\mathbf{T}(r), \mathbf{D}\{\mathbf{T}(r)\}] | \mathbf{D}(0) = \underline{d}] &= X[r, \underline{d}] \{M_S(\underline{d})\}^r \\ &+ \sum_{n=1}^r \sum_{j=1}^S E[X[\mathbf{T}(r) + n, \mathbf{D}\{\mathbf{T}(r)\}] | \mathbf{D}(0) = \underline{d} - \mathbf{1}^j] \\ &\times \{M_S(\underline{d})\}^{n-1} \{M_{j-1}(\underline{d}) - M_j(\underline{d})\}, \end{aligned} \quad (4.1)$$

where

$$\begin{aligned} M_j(\underline{d}) &= E_\varepsilon \left[\left\{ \prod_{i=1}^j \tilde{q}_i(d_i) \right\}^\varepsilon \right] \\ &= E_\varepsilon \left[\exp \left\{ -\varepsilon \left\{ \sum_{i=1}^j -\ln \tilde{q}_i(d_i) \right\} \right\} \right], \quad 1 \leq j \leq S. \end{aligned} \quad (4.2)$$

The latter expressions may be evaluated in terms of the Laplace transforms of the ε -distribution. To obtain the mean number of tests with a run of successes, take $X[\mathbf{T}, \mathbf{D}] = \mathbf{T}$ in Eq. (4.1).

It may also make sense to introduce an i.i.d. sequence $\varepsilon_1(f), \varepsilon_2(f), \dots$, of random variables relating to environments encountered when deploying the system in the field. Any such sequence is deemed independent of the ε -sequence above and, most significantly, may be differently distributed. If we now wished to compute the mean probability of system mission survival in the field after a “first run of r ” test, we then take

$$\begin{aligned} X[\mathbf{T}, \mathbf{D}] &= E_{\varepsilon(f)} \left[\left(\prod_{i=1}^S \tilde{Q}_i(\mathbf{D}_i) \right)^{\varepsilon(f)} \right] \\ &= E_{\varepsilon(f)} \left[\exp \left\{ -\varepsilon(f) \left(\sum_{i=1}^S -\ln \tilde{Q}_i(\mathbf{D}_i) \right) \right\} \right] \end{aligned}$$

and use Eq. (4.1). The following figures display the important role that the presence of environmental variability may play in the ability of operational testing to result in the fielding of reliable systems:

Figure 8 displays probabilities of system field success for a system that has been tested until there is a run of five successes. The testing environmental random variables have a gamma distribution with mean 1 and shape parameter 0.5. The field

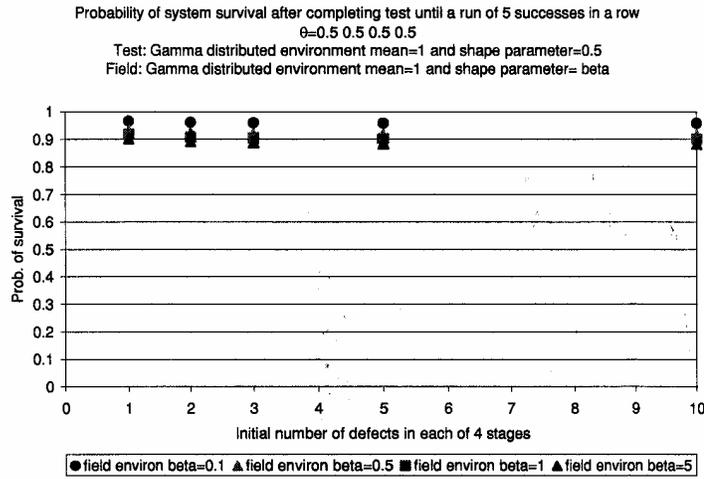


Fig. 8.

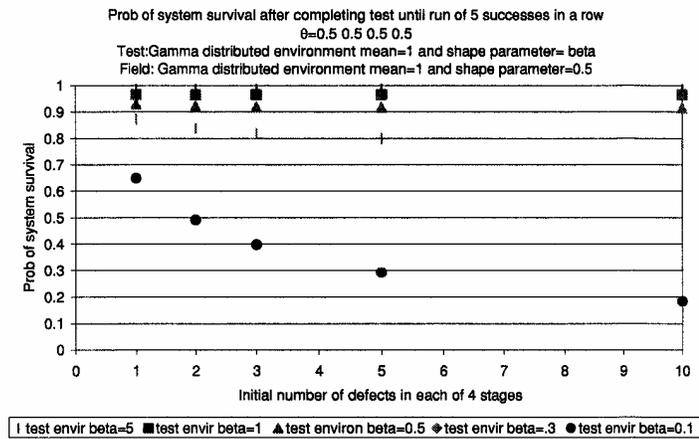


Fig. 9.

environmental random variable has a gamma distribution with mean 1 and shape parameter β , which has been made widely variable. Note that the smaller β is, the greater is the probability of field system survival. In the present case, the testing environment is variable enough to produce an effect that is, in quite disparate field conditions, fairly insensitive to the distribution of random field effects.

As in Fig. 8, Fig. 9 (respectively 10) displays probabilities of field success (respectively, the number of tests required) for a system that is to be tested until there is a run of 5 successes in a row. Here the field environmental random variable has a gamma distribution with mean 1 and shape parameter equal to 0.5. The test environment random variables have a gamma distribution with mean 1 and variable

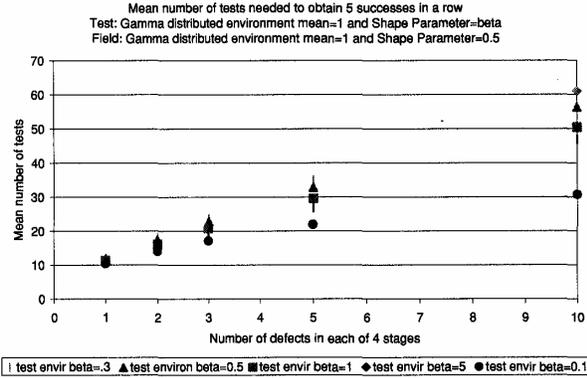


Fig. 10.

shape parameter β . The small shape parameter, $\beta = 0.1$, results in smaller mean number of tests required, but at the price of a smaller probability of field success. This is because a gamma density function with $\beta = 0.1$ has most of its mass close to 0. Thus, most of the time the probability that a defect is revealed during a test is close to 0, and the test is over too soon to eliminate many defects. However, since the field environment random variable has a shape parameter equal to 0.5, the defects remaining after the test is completed are likely to be triggered in the field.

Variable test environments that allow a disproportionate number of excessively benign environments, even though balanced by some that are excessively stringent, can thus severely bias the quality of the fielded product.

5. Bayesian Formulations

A natural approach to the uncertainty concerning the numbers of design defects initially present is a Bayesian one in which $\mathbf{D}_i(0)$ is treated as a random variable with (prior) distribution $\Pi^i = \{\Pi_d^i, d \geq 0\}$, $1 \leq i \leq S$. In what follows, we shall suppose that the random variables $\mathbf{D}_i(0)$, $1 \leq i \leq S$ are independent and that the conditional model for failure discovery and removal is as in Sec. 3, with $\tilde{q}_i(d_i)$ the conditional probability of subsystem i success, given d_i defects present. In such a setup, consider the situation following t tests of the system.

Each subsystem i will have its own history $H_{it} = \{x_{i1}, x_{i2}, \dots, x_{it}\}$ where each x_{ij} takes one of three possible values, namely

- {subsystem i was not subjected to scrutiny during test j because of the failure of an earlier subsystem} $\equiv O_{ij}$;
- {subsystem i was subjected to scrutiny during test j and operated successfully} $\equiv S_{ij}$; and

{subsystem i was subjected to scrutiny during test j and a defect was activated and removed} $\equiv F_{ij}$.

The first of these cannot occur for subsystem 1. Let $\Pi^i(t)$ be the inferred (posterior) distribution of $\mathbf{D}_i(t)$ upon suitable application of Bayes' theorem. Updating is described as follows:

$$\Pi_d^i(t+1) \begin{cases} = \Pi_d^i(t), & \text{if } x_{it+1} = O_{it+1}; \\ \propto \Pi_d^i(t) \bar{q}_i(d), & \text{if } x_{it+1} = S_{it+1}; \text{ and} \\ \propto \Pi_{d+1}^i(t) \{1 - \bar{q}_i(d+1)\}, & \text{if } x_{it+1} = F_{it+1}. \end{cases} \quad (5.1)$$

In general, the posterior $\Pi^i(t)$ will depend upon the entire history H_{it} , and in particular will depend upon the order in which successes and failures occur.

5.1. Conjugate families of prior distributions

In this highly complex scenario it seems reasonable to make an initial search for simplicity in the form of the elucidation of conjugate structures. To the authors' knowledge, the only families of conjugate priors for this problem arise from a requirement that each posterior $\Pi^i(t)$ should depend upon the history H_{it} only through $\{\mathbf{S}_i(t), \mathbf{F}_i(t)\}$, where $\mathbf{S}_i(t)$ is the total number of successful operations of subsystem i during t tests and $\mathbf{F}_i(t)$ is similarly defined in terms of failures, (i.e., defect activations and removals). Work by Benkherouf and Bather⁵ in the context of oil exploration implies that this requirement forces a binomial conditional model of the form

$$\bar{q}_i(d) = \theta_i^d, d \geq 0, \text{ for some } \theta_i, 0 < \theta_i < 1, \quad (5.2)$$

as discussed in Sec. 3. With this conditional model in place, a family of prior distributions introduced by Glazebrook²¹ is conjugate for this problem. Let $\mathbf{D}_i(0) \sim \Pi^i(0) \equiv \Pi(\lambda_i, \theta_i, \phi_i)$, $1 \leq i \leq S$, where the probability mass function (p.m.f.) for the conjugate prior $\Pi(\lambda, \theta, \phi)$ is given by

$$\Pi_d(\lambda, \theta, \phi) = \Pi_0(\lambda, \theta, \phi) \lambda^d \theta^{\phi d(d-1)/2} \left\{ \prod_{t=1}^d (1 - \theta^t) \right\}^{-1}, \quad d \geq 0 \quad (5.3)$$

where $\Pi_0(\lambda, \theta, \phi)$ is a normalizing constant. The parameter space associated with this family is $\{0 < \lambda < 1, 0 < \theta < 1, \phi = 0\} \cup \{\lambda > 0, 0 < \theta < 1, \phi > 0\}$. The first parameter λ may be interpreted as an overall rate of finding failures, while ϕ may be thought of as a rate of depletion of defects in a subsystem under failure, and subsequent defect removal. The parameter θ is always assigned the value of the probability in the conditional binomial model in Eq. (5.2). The prior family in Eq. (5.3) is quite versatile. For example, when setting $\phi = 0$ there are regions of the parameter space in which the distribution may approach either a Poisson or a geometric distribution.

With the prior $\Pi(\lambda_i, \theta_i, \phi_i, \cdot)$ in Eq. (5.3) and the conditional model (5.2), the posterior distribution for $\mathbf{D}_i(t)$ is given (utilizing Eq. (5.1)) by

$$\Pi_d^i(t) \equiv \Pi_d\{\lambda_i \theta_i^{\mathbf{S}_i(t) + \phi_i \mathbf{F}_i(t)}, \theta_i, \phi_i\}, \quad 1 \leq i \leq S, \quad (5.4)$$

for $d = 0, 1, 2, \dots$. We conclude from Eq. (5.4) that $\{\mathbf{S}_i(t) + \phi_i \mathbf{F}_i(t)\}$, $1 \leq i \leq S$, are sufficient statistics for the problem. This, however, raises a serious issue regarding the applicability of these conjugate structures in the current context. Recall the role of the first parameter as an overall rate of finding failures. This in turn implies that a natural proposal would be to stop testing after t tests when the values

$$\lambda_i \theta_i^{\mathbf{S}_i(t) + \phi_i \mathbf{F}_i(t)}, \quad 1 \leq i \leq S \quad (5.5)$$

are sufficiently small. Indeed, Bayes optimal tests with reasonable cost criteria would be expected to have such a form. However, this observation raises the possibility that a testing program could be stopped under such a rule following a test which resulted in system failure. Our judgment is that no such test could be credible within our envisaged field of applications. While the family of priors in Eq. (5.3) alone is versatile, the combination of these priors together with the conditional model Eq. (5.2) (which guarantees conjugacy) is not an acceptable basis for the development of credible test designs and prior/posterior analyses.

5.2. Stagewise overvariability

Section 3 argues for a modification of Eq. (5.2) by the representation of extra-binomial stage-to-stage variability in the conditional model. Hence, we now consider replacing Eq. (5.2) by

$$\tilde{q}_i(d) = E(\theta^d) \quad \text{with } \theta \sim G_i, \quad 1 \leq i \leq S. \quad (5.6)$$

Recall that the binomial model (5.2) was required for the simple structures above based upon sufficient statistics (\mathbf{S}, \mathbf{F}) . We conclude that, with the more general Eq. (5.6), the posterior distribution $\Pi^i(t)$ will depend upon the entire history $H_{it} = \{x_{i1}, x_{i2}, \dots, x_{it}\}$, excepting only those entries x_{ij} equal to O_{ij} . Put another way, $\Pi^i(t)$ will depend upon the complete sequence of successes and failures to date. It emerges that, while we lose simplicity of structure by generalizing in this way, we make important advances in applicability of the model, and in addition, develop a rationale for run tests as good stopping criteria.

We focus first on a single subsystem and, for the present, drop subsystem identifier i . The subsystem has $\mathbf{D}(0)$ defects initially with associated (prior) distribution $\Pi = \{\Pi_d, d \geq 0\}$. The conditional model is $\tilde{q}(d)$, with $\tilde{p}(d) = 1 - \tilde{q}(d)$, $d \geq 0$. We consider a sequence of t tests during which the subsystem is subject to scrutiny upon $\varphi + \sum_{j=1}^{\varphi+1} \sigma_j$ occasions of which φ result in failure (and defect removal) and $\sum_{j=1}^{\varphi+1} \sigma_j$ result in system success. More precisely, σ_1 is the number of successes before the first subsystem failure, $\sigma_{\varphi+1}$ is the number of successes following the last φ th subsystem failure and σ_j , $2 \leq j \leq \varphi$, is the number of successes between failures $(j-1)$ and j . We write $\{\sigma_1, \sigma_2, \dots, \sigma_{\varphi+1}, \varphi\}$ for this data configuration.

By repeated application of Eq. (5.1), the posterior probability of d subsystem defects remaining following these t tests is given by

$$\begin{aligned} \Pi_d(t) &\equiv \Pi_{d|\sigma_1, \sigma_2, \dots, \sigma_{\varphi+1}, \varphi} \\ &\propto \Pi_{d+\varphi} \left\{ \prod_{j=1}^{\varphi} \tilde{p}(d+j) \right\} \left[\prod_{k=1}^{\varphi+1} \{\tilde{q}(d+\varphi+1-k)\}^{\sigma_k} \right]. \end{aligned} \quad (5.7)$$

Further simplification results in the special case $\tilde{q}(d) = 1/(d+1)$ which results from taking $\theta \sim U[0, 1]$ for all stages in Eq. (5.6); see Subsec. (3.1). The posterior distribution in Eq. (5.7) then becomes

$$\Pi_{d|\sigma_1, \sigma_2, \dots, \sigma_{\varphi+1}, \varphi} \propto \Pi_{d+\varphi} \left(\frac{d+1}{d+\varphi+1} \right) \left\{ \prod_{k=1}^{\varphi+1} \left(\frac{1}{d+\varphi+2-k} \right)^{\sigma_k} \right\}. \quad (5.8)$$

We now perform some calculations, which shed light upon the nature of updating and reliability growth in this general context. A key focus of the analysis will concern how the posterior probability of system survival in the field varies with the data. When we discuss the full system we shall need to restore the subsystem identifier i . Consider now two subsystem data configurations $\{\sigma_1, \sigma_2, \dots, \sigma_{\varphi+1}, \varphi\} \equiv \{\sigma, \varphi\}$ and $\{\sigma'_1, \sigma'_2, \dots, \sigma'_{\varphi+1}, \varphi\} \equiv \{\sigma', \varphi\}$.

Definition. Data configuration $\{\sigma, \varphi\}$ dominates configuration $\{\sigma', \varphi\}$ if $\sum_{k=1}^j \sigma_k \leq \sum_{k=1}^j \sigma'_k$, $1 \leq j \leq \varphi$ and $\sum_{k=1}^{\varphi+1} \sigma_k = \sum_{k=1}^{\varphi+1} \sigma'_k$.

The above definition is describing a partial ordering between data configurations in which the dominating sequence has the same (total) number of successes and failures, but has the failures earlier. Note that in the models based on the binomial conditional model in Eq. (5.2) with conjugate prior, the posterior distributions for the two sequences would be identical. This is no longer the case.

Let $Q_i(\sigma^i, \varphi^i)$ for the predictive probability of field mission success for subsystem i following data (σ^i, φ^i) , $1 \leq i \leq S$, where

$$Q_i\{(\sigma^i, \varphi^i)\} = \sum_{d \geq 0} \Pi_{d|\sigma^i, \varphi^i}^i \tilde{Q}_i(d), \quad 1 \leq i \leq S. \quad (5.9)$$

The corresponding predictive probability for the system as a whole is

$$Q\{(\sigma^1, \varphi^1), (\sigma^2, \varphi^2), \dots, (\sigma^S, \varphi^S)\} = \prod_{i=1}^S Q_i\{(\sigma^i, \varphi^i)\}. \quad (5.10)$$

In the following result we use $\pi_{\sigma^i, \varphi^i}^i$ as a notation shorthand for the (posterior) distribution for the number of defects in subsystem i following data configuration (σ^i, φ^i) , $1 \leq i \leq S$.

Theorem 2. For any choices of prior distribution Π^i with $\Pi_d^i \geq 0$, $d \in N$, and conditional model (5.6) for which $\tilde{q}_i(1) = E_{G_i}(\theta) \in (0, 1)$, $1 \leq i \leq S$, the following hold:

- (1) If $\{\sigma^i, \varphi^i\}$ dominates $\{\sigma'^i, \varphi^i\}$ then $\Pi_{\sigma^i, \varphi^i}^i$ is stochastically smaller than $\Pi_{\sigma'^i, \varphi^i}^i$;
(2) If the sequence $\{\tilde{Q}_i(d), d \geq 0\}$ is decreasing and $\{\sigma^i, \varphi^i\}$ dominates $\{\sigma'^i, \varphi^i\}$, $1 \leq i \leq S$, then

$$Q_i\{\{\sigma^i, \varphi^i\}\} \geq Q_i\{\{\sigma'^i, \varphi^i\}\}, \quad 1 \leq i \leq S,$$

and hence

$$Q\{(\sigma^1, \varphi^1), (\sigma^2, \varphi^2), \dots, (\sigma^S, \varphi^S)\} \geq Q\{(\sigma'^1, \varphi^1), (\sigma'^2, \varphi^2), \dots, (\sigma'^S, \varphi^S)\};$$

- (3) If the sequence $\{\tilde{Q}_i(d), d \geq 0\}$ is decreasing, then $Q_i(\sigma^i, \varphi^i)$ is non-decreasing in each σ_j^i , $1 \leq j \leq \varphi^i + 1$, $1 \leq i \leq S$, as is $Q\{(\sigma^1, \varphi^1), (\sigma^2, \varphi^2), \dots, (\sigma^S, \varphi^S)\}$.

Proof. (1) Let j be such that $\sigma_j'^i \geq 1$ and $1 \leq j \leq \varphi^i$. Consider configuration $\{\sigma'^i + \underline{1}^{j+1} - \underline{1}^j, \varphi^i\}$. Direct application of Eq. (5.7) shows that

$$\begin{aligned} \frac{\Pi_{d|\sigma^i, \varphi^i}^i}{\Pi_{d|\sigma'^i + \underline{1}^{j+1} - \underline{1}^j, \varphi^i}^i} &= K_i(\sigma'^i, \varphi^i, j) \frac{\tilde{q}_i(d + \varphi^i + 1 - j)}{\tilde{q}_i(d + \varphi^i - j)} \\ &= K_i(\sigma'^i, \varphi^i, j) \frac{E_{G_i}(\theta^{d + \varphi^i + 1 - j} | \varphi^i)}{E_{G_i}(\theta^{d + \varphi^i - j} | \varphi^i)}, \quad d \geq 0, \end{aligned} \quad (5.11)$$

for some constant $K_i(\sigma'^i, \varphi^i, j)$. But since distribution G_i has support contained in $[0, 1]$ it is straightforward, taking limits of discrete distributions, to show that $\{E_{G_i}(\theta^{d+1})/E_{G_i}(\theta^d), d \geq 0\}$ is a non-decreasing sequence. It follows immediately from Eq. (5.11) that the distribution $\Pi_{\sigma'^i + \underline{1}^{j+1} - \underline{1}^j, \varphi^i}^i$ is smaller than $\Pi_{\sigma'^i, \varphi^i}^i$ in the likelihood ratio ordering, and hence also in the stochastic ordering. However, we can move from (σ'^i, φ^i) to dominating configuration (σ^i, φ^i) by means of a sequence of transitions of the form $(\tilde{\sigma}^i, \varphi^i) \rightarrow (\tilde{\sigma}^i + \underline{1}^{j+1} - \underline{1}^j, \varphi^i)$ for some j . This, together with the transitivity of stochastic ordering yields (1).

(2) is a simple consequence of (1).

The proof of (3) involves straightforward application of Eq. (5.7) and is omitted.

We shall now infer from Theorem 1 two further results, both of which are strongly suggestive of the appropriateness of the runs tests as effective designs in the current context. To this end we introduce $C(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$ as the equivalence class of those data configurations for the system with numbers of successes and failures in the respective subsystems summarized by vectors $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{F}}$. Hence, for example, $\tilde{\mathbf{S}}, \tilde{\mathbf{F}}$ together summarize a situation in which subsystem j has passed test \tilde{S}_j times but failed on \tilde{F}_j occasions, $1 \leq j \leq S$. Observe that, for vectors $(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$ to be feasible, for our experimental context we must have

$$\tilde{S}_j = \tilde{S}_{j+1} + \tilde{F}_{j+1}, \quad 1 \leq j \leq S - 1.$$

We write

$$\underline{\sigma}, \underline{\varphi} = \{(\sigma^1, \varphi^1), (\sigma^2, \varphi^2), \dots, (\sigma^S, \varphi^S)\} \in C(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$$

where

$$\varphi^j = \tilde{\mathbf{F}}_j, \sum_{i=1}^{\tilde{\mathbf{F}}_j+1} \sigma_i^j = \tilde{\mathbf{S}}_j, \quad 1 \leq j \leq S.$$

We shall use the term *successes earliest* to identify that member of each $C(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$ for which the successes in each stage come before any failures; that is,

$$\sigma_1^j = \tilde{\mathbf{S}}_j; \quad \sigma_i^j = 0, \quad 2 \leq i \leq \tilde{\mathbf{F}}_j + 1; \quad 1 \leq j \leq S.$$

The term *successes latest* identifies that member of each $C(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$ for which all the successes come at the end of the test; that is,

$$\sigma_i^j = 0, \quad 1 \leq i \leq \tilde{\mathbf{F}}_j; \quad \sigma_{\tilde{\mathbf{F}}_j+1}^j = \tilde{\mathbf{S}}_j; \quad 1 \leq j \leq S.$$

Note that the *successes earliest* configuration reflects experimental results that begin testing with a run of $\tilde{\mathbf{S}}_S$ system successes while the *successes latest* configuration ends testing with a run of $\tilde{\mathbf{S}}_S$ system successes. \square

Theorem 3. *If the sequence $\{\tilde{Q}_i(d), d \geq 0\}$ is decreasing, $1 \leq i \leq S$ then the predictive probability of field mission success, $Q\{\underline{\sigma}, \underline{\varphi}\}$, is maximized over $\{\underline{\sigma}, \underline{\varphi}\} \in C(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$ by the successes latest configuration and minimized by the successes earliest configuration, for all feasible $(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$.*

Proof. We observe that the *successes latest* member of $C(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$ dominates all other members, while the *successes earliest* member is dominated by all other members. The results is then an immediate consequence of Theorem 2(2). \square

Comment. We note that in the models based on the binomial model in Eq. (5.2) all members of $C(\tilde{\mathbf{S}}, \tilde{\mathbf{F}})$ have the same predictive probability of field mission success irrespective of the disposition of successes and failures.

Theorem 4. *If $\Pi_d^i > 0, 0 \leq d \leq D_i$ for some $D_i \geq 0, 1 \leq i \leq S$, then during a run of r system successes, the predictive probability of field mission success approaches 1 at a geometric rate in the number of system successes to date.*

Proof. Following a run of r successes, each subsystem data configuration from the point at which the run starts is $(r, 0)$. We use $\hat{\Pi}^i$ for the posterior for subsystem i , which is current at the start of the run. It is easy to show from Eq. (5.7) that the condition on the prior Π^i in the statement of the theorem guarantees that $\hat{\Pi}_0^i > 0$. By Eq. (5.6) we have that, in an obvious notation, the posterior probability of field mission success after a run of r successes is

$$\hat{Q}_i\{(r, 0)\} \geq \hat{\Pi}_0^i|_{r,0} = \frac{\hat{\Pi}_0^i}{\sum_{d \geq 0} \hat{\Pi}_d^i \{\tilde{q}_i(d)\}^r} \geq \frac{\hat{\Pi}_0^i}{\hat{\Pi}_0^i + \{1 - \hat{\Pi}_0^i\} \{\tilde{q}_i(1)\}^r}, \quad (5.12)$$

where inequality (5.12) utilizes the decreasing nature of the sequence $\{E_{G_i}(\theta^d), d \geq 0\} \equiv \{\tilde{q}_i(d), d \geq 0\}$. From Eq. (5.12) we deduce that the predictive probability of field mission success for the whole system following the run of r successes is

$$\begin{aligned} \hat{Q}\{(r, 0)\} &= \prod_{i=1}^S \hat{Q}_i\{(r, 0)\} \geq \prod_{i=1}^S \left[1 + \frac{(1 - \hat{\Pi}_0^i)}{\hat{\Pi}_0^i} \{\tilde{q}_i(1)\}^r \right]^{-1} \\ &\geq 2 - \prod_{i=1}^S \left[1 + \frac{(1 - \hat{\Pi}_0^i)}{\hat{\Pi}_0^i} \{\tilde{q}_i(1)\}^r \right] \\ &\geq 1 - \left(\left[1 + \sum_{i=1}^S \frac{(1 - \hat{\Pi}_0^i)}{\hat{\Pi}_0^i} \frac{\{\tilde{q}_i(1)\}^r}{S} \right]^S - 1 \right) \end{aligned} \quad (5.13)$$

using the inequality between the geometric mean and the arithmetic mean. The result is a straightforward consequence. \square

Comment. We point out that the condition on the prior in Theorem 4 is a perfectly natural one and simply prohibits giving zero prior probability to some number of defects, d say, in some subsystem i while giving positive probability to values above and below d .

The choice of the length of a run required, r , for a runs test may be assessed by means of a prior analysis focusing on such key measures as the mean probability of system survival of a mission in the field following the test; the mean time to the conclusion of testing; and the probability that the field probability of mission success at the end of testing is greater than $1 - \alpha_1$.

Suppose, for example, that we wish to select r to maximize an objective

$$E_{\Pi}\{-c\mathbf{T} + \tilde{Q}_{\mathbf{T}}\} \quad (5.14)$$

where Π is the prior chosen, \mathbf{T} is the number of tests performed, c is a (suitably standardized) cost per test and

$$\tilde{Q}_{\mathbf{T}} = \begin{cases} 1, & \text{for field mission success following } \mathbf{T} \text{ tests} \\ 0, & \text{otherwise.} \end{cases}$$

The quantity in Eq. (5.14) can be calculated using the computations described in Sec. 3.3 and is given by

$$\sum_{\underline{d}} \left\{ \prod_{i=1}^S \Pi^i(d_i) \right\} \{-c\tau_r(\underline{d}) + p_r(\underline{d})\} \quad (5.15)$$

for the “ r runs” test. A Bayes optimal design within the class of runs tests is obtained by choosing r to maximize the quantity in Eq. (5.15).

An alternative prior approach to choosing a runs test would choose the smallest r to guarantee that the prior-posterior probability of field mission success (following the test) exceeding some value $1 - \alpha_1$, say, is at least some value $1 - \alpha_2$. To describe

how this might be achieved, we expand the notation in Eq. (3.9) to $\gamma_r(\underline{a}|\underline{d})$ to reflect a conditioning on the initial state \underline{d} . We write

$$\Delta_{1-\alpha_1} = \left\{ \underline{d} : \prod_{i=1}^S \tilde{Q}_i(d_i) \geq 1 - \alpha_1 \right\}$$

for the set of defect configurations with a corresponding field probability of mission success at least $1 - \alpha_1$. The identified design criterion is achieved by identifying the smallest r with which

$$\sum_{\underline{a} \in \Delta_{1-\alpha_1}} \sum_{\underline{d}} \left\{ \prod_{i=1}^S \Pi^i(d_i) \right\} \gamma_r(\underline{a}|\underline{d}) \geq 1 - \alpha_2.$$

Following an r runs test, a posterior probability may focus on, for example, the construction of a Bayes confidence interval for the probability of field success. To achieve this, list the sequence of defect configurations in descending order of $\prod_{i=1}^S \tilde{Q}_i(d_i)$, the probability of field mission success with the remaining defects d_i , $i = 1, \dots, S$. We denote the resulting sequence $\underline{d}^{(1)} = 0, \underline{d}^{(2)}, \underline{d}^{(3)}$, and so on. Let $N(r, \alpha, D)$ be the smallest n for which the posterior analysis of the collection $\{\underline{d}^{(1)}, \underline{d}^{(2)}, \dots, \underline{d}^{(n)}\}$ is at least $1 - \alpha$, where D denotes the test data. It then follows that $[1, \prod_{i=1}^S \tilde{Q}_i\{d_i^{N(r, \alpha, D)}\}]$ is a Bayes $(1 - \alpha)$ -confidence interval for the probability of field mission success.

5.3. Numerical example

Table 1 reports results from a numerical study of the probability of system field success after a test, which ends with the first occurrence of r successes in a row. The study elucidates the fact that any prior analysis to determine the test design (i.e., choice of r) is likely to be significantly impacted by our beliefs concerning whether defects are more numerous in the earlier or later stages of the system. The system consists of four stages. Given d_s defects in stage s , $s = 1, \dots, 4$, the conditional probability that the system passes one test is $\prod_{s=1}^4 q_s(d_s)$ where

$$q_s(d_s) = E[\theta_s^{d_s}] = \frac{\Gamma(a_s + b_s)}{\Gamma(b_s)} \frac{\Gamma(b_s + d_s)}{\Gamma(a_s + b_s + d_s)}$$

with θ_s having a beta distribution. Two cases of randomized θ_s are considered. In case A, each θ is drawn from a uniform distribution on $[0, 1]$ independently for each stage and test. In case B, θ_s is drawn from a beta distribution with mean $b_s/(a_s + b_s)$ for

$$(a_s, b_s) = \begin{cases} (.9, .1) & \text{for } s = 1, \\ (.7, .3) & \text{for } s = 2, \\ (.3, .7) & \text{for } s = 3, \\ (.1, .9) & \text{for } s = 4. \end{cases}$$

Table 1. Mean of summary statistics for simulations of testing until obtain a run of r successes.

Mean initial# defects stage 1	Mean initial# defects stage 2	Mean initial# defects stage 3	Mean initial# defects stage 4	# repl	Mean of mean prob field surv	Mean prob that the prob field surv ≥ 7	Mean prob that the prob field surv ≥ 8	Mean prob that the prob field surv ≥ 9	Mean prob that the prob field surv ≥ 95	# suc in a row: r	prob surv in field for each remaining defect	over-var prob of surv during test	Mean of the mean number of tests needed to obtain r successes in a row
2.75	2.75	2.75	2.75	25	0.96 (0.00)	0.96 (0.00)	0.96 (0.00)	0.84 (0.00)	0.84 (0.00)	3	0.8	A	14.46 (0.49)
2.75	2.75	2.75	2.75	25	0.59 (0.03)	0.32 (0.05)	0.32 (0.05)	0.13 (0.02)	0.13 (0.02)	3	0.8	B	14.1 (0.64)
2.75	2.75	2.75	2.75	25	0.74 (0.02)	0.56 (0.05)	0.56 (0.05)	0.31 (0.04)	0.31 (0.04)	5	0.8	B	17.67 (0.55)
2.75	2.75	2.75	2.75	25	0.82 (0.02)	0.70 (0.04)	0.70 (0.04)	0.44 (0.04)	0.44 (0.04)	7	0.8	B	21.85 (0.61)
2.75	2.75	2.75	2.75	25	0.91 (0.01)	0.96 (0.00)	0.94 (0.01)	0.69 (0.04)	0.44 (0.04)	7	field prob same as testing	B	21.85 (0.61)
1	2	3	5	25	0.95 (0.00)	0.95 (0.00)	0.95 (0.00)	0.83 (0.00)	0.83 (0.00)	3	0.8	A	15.09 (0.7)
1	2	3	5	25	0.43 (0.03)	0.12 (0.02)	0.12 (0.02)	0.04 (0.01)	0.04 (0.01)	3	0.8	B	11.31 (0.49)
1	2	3	5	25	0.56 (0.03)	0.28 (0.02)	0.28 (0.02)	0.13 (0.01)	0.13 (0.01)	5	0.8	B	23 (0.65)
1	2	3	5	25	0.66 (0.03)	0.44 (0.04)	0.44 (0.04)	0.24 (0.02)	0.24 (0.02)	7	0.8	B	23 (0.82)
1	2	3	5	25	0.87 (0.01)	0.97 (0.00)	0.82 (0.04)	0.44 (0.04)	0.24 (0.02)	7	field prob same as testing	B	23 (0.82)
5	3	2	1	25	0.96 (0.00)	0.97 (0.00)	0.97 (0.00)	0.85 (0.00)	0.85 (0.00)	3	0.8	A	14.19 (0.72)
5	3	2	1	25	0.80 (0.02)	0.69 (0.05)	0.69 (0.05)	0.38 (0.05)	0.38 (0.05)	3	0.8	B	13.31 (0.67)
5	3	2	1	25	0.88 (0.02)	0.84 (0.04)	0.84 (0.04)	0.56 (0.05)	0.56 (0.05)	5	0.8	B	17.02 (0.84)
5	3	2	1	25	0.92 (0.01)	0.90 (0.03)	0.90 (0.03)	0.68 (0.05)	0.68 (0.05)	7	0.8	B	20.22 (0.95)
5	3	2	1	25	0.95 (0.01)	0.97 (0.00)	0.94 (0.01)	0.87 (0.02)	0.68 (0.05)	7	field prob same as testing	B	23 (0.95)

In all but three cases, the field probability of system success is $\prod_{s=1}^4 0.8^{d_s(r)}$, where $d_s(r)$ is the number of defects remaining in stage s after the test is complete. The initial numbers of defects in each stage are independently drawn from Poisson (prior) distributions with the means noted in the table. There are 25 replications for each case. Displayed are the mean of the expected probability of system field mission success, together with the means of the probabilities that the probability system field mission success after the test is greater than or equal to 0.7, 0.8, 0.9, and 0.95. These values estimate the quantities

$$\sum_{\underline{a} \in \Delta_{1-\alpha}} \sum_{\underline{d}} \left\{ \prod_{i=1}^S \Pi^i(d_i) \right\} \gamma_r(\underline{a}|\underline{d})$$

for $r = 3, 5$ and 7 and $\alpha = 0.3, 0.2, 0.1$ and 0.05 . We also record the average of the expected number of tests required to obtain a run of r successes; this estimates the quantity

$$\sum_{\underline{d}} \left\{ \prod_{i=1}^S \Pi^i(d_i) \right\} \tau_r(\underline{d}).$$

The standard errors appear in parentheses.

The distribution of the $\theta_s, s = 1, \dots, 4$ has a great effect on the probability of successful field performance after the test. In case B, the defects in stage 4 are less likely to reveal themselves during the test. Thus for case B, the probability of field success after a test until a run of three successes is smaller than for the case of uniformly distributed $\theta_s, s = 1, \dots, 4$.

The initial mean number of defects in each stage also affects the probability of field success. The case with a mean of five defects in stage 4 has the smallest mean of the expected probability of field success after a test. The mean of the expected probabilities of field success after a test until there is a run of seven successes in a row is 0.66 for this case.

The mean of the expected number of tests needed to obtain a run of r successes for the cases displayed is somewhat insensitive to the pattern of the initial mean number of defects in each stage, and the probability of defect discovery during test.

In all but three of the cases the probability of a defect in a stage causing failure during use in the field is 0.8, which is different than these probabilities during testing. In the three cases in which the probability of field success is the same as that in testing, the mean expected probabilities of field success are higher. It is important to design tests so that they represent field conditions as closely as possible.

6. Summary and Conclusions

In this paper we consider models of overall system testing to achieve reliability growth by design defect identification and removal. This is sometimes referred to

as *Test-and-Fix* (TAF). We consider a system with S stages in sequence; if a test reveals a defect in stage s , the later stages $s + 1, \dots, S$ are not subjected to the test. We assume that, at most, one defect is removed per test.

A sequential test plan that ensures that all the stages will be tested at least r times is to test until there is a *run of r consecutive system successes*. A system success means that all the stages operate successfully during the test, which implies that the propensities to fail of remaining design defects is likely to be small. Results obtained for a Bayesian model formulation suggest that, while not being Bayes optimal in a formal sense, a runs test provides a simple and effective test stopping rule for a range of reasonable cost criteria.

We obtain analytical procedures to calculate the expected probability of field system mission success after successful completion of a runs test, the distribution of the probability of system field mission success after a successful runs test, and the expected number of individual system tests required to achieve a run of r successes, and hence test termination. Numerical studies indicate that the probability of system field success after a runs test can be quite sensitive to the probabilities that a test activates defects in each of the stages. However, the mean number of tests required to obtain a run of r successful tests appears to be relatively insensitive to these activation probabilities.

Acknowledgments

We thank Professor Bent Natwig, University of Oslo, for his careful reading of the manuscript.

References

1. H. Ascher and H. Feingold, *Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes, Lecture Notes in Statistics 7* (Marcel Dekker, New York, 1984).
2. H. S. Balaban, "Reliability growth models," *Journal of Environmental Sciences* **21** (1978), pp. 11–18.
3. R. E. Barlow and E. M. Scheuer, "Reliability growth during a development testing program," *Technometrics*, **8**, (1966), pp. 53–60.
4. D. R. Barr, "A class of general reliability growth models," *Operations Research* **18** (1970), pp. 52–65.
5. L. Benkherouf and J. A. Bather, "Oil exploration: sequential decisions in the face of uncertainty," *Journal of Applied Probability* **25** (1988), pp. 529–543.
6. L. Benkherouf, K. D. Glazebrook and R. W. Owen, "Gittins indices and oil exploration," *Journal of the Royal Statistical Society* **B54** (1992), pp. 229–241.
7. G. K. Bhattacharyya, A. Fries and R. A. Johnson, "Properties of continuous analog estimators for a discrete reliability-growth model," *IEEE Transactions on Reliability* **38** (1989), pp. 373–378.
8. H. W. Block and T. H. Savits, "Burn-In," *Statistical Science* **12/1** (1997), pp. 1–13.
9. V. Bogdonavicius and M. S. Nikulin, "Estimation in Degradation Models with Explanatory Variables. Report, Centre des Consultations Opti-Stats, Responsable

- de L'Equipe d'Accueil" Statistique mathématique et ses Applications, Univ. Victor Segalen, Bordeaux, France (2000).
10. R. Calabria, M. Guida and G. Pulcini, "A Bayes procedure for estimation of current system reliability," *IEEE Transactions on Reliability* **41** (1992), pp. 616–621.
 11. H. Chernoff and S. N. Ray, "A Bayes sequential sampling inspection plan," *Annals of Mathematical Statistics* **36/5** (1965).
 12. H. Chernoff, "Sequential models for clinical trials," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, CA, 1966).
 13. R. G. Easterling, M. Mazumdar, F. W. Spencer, and K. V. Diegert, "System-based component-test plans and operating characteristics: binomial data," *Technometrics* **33/3** (1991), pp. 287–298.
 14. W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume II (John Wiley & Sons, Inc., New York, 1966).
 15. A. Fries and L. M. Maillart, "TAAF (Test-Analyze-And-Fix) Stopping Rules for Maximizing the Operational Utility of One-Shot Systems," IDA Document D-1880. Institute of Defense Analyses, Arlington, VA, (1996).
 16. A. Fries, "Discrete reliability-growth models based on a learning-curve property," *IEEE Transactions on Reliability* **42** (1993), pp. 303–306.
 17. D. P. Gaver and P. A. Jacobs, "Methodology for an operationally-based test length decision," *IEEE Transactions* **30** (1998), pp. 1129–1134.
 18. D. P. Gaver and P. A. Jacobs, "Testing or fault-finding for reliability growth: A missile destructive-test example," *Naval Research Logistics* **44** (1997), pp. 623–637.
 19. D. P. Gaver and P. A. Jacobs, "WEBSIGHT: An internet tool for uncertainty analysis in operational testing," *American Statistical Association Joint Annual Meetings* (Baltimore, MD, August 8–12, 1999).
 20. J. C. Gittins, *Multi-armed Bandit Allocation Indices* (John Wiley & Sons, Inc., New York, 1989).
 21. K. D. Glazebrook, "On a family of prior distributions for a class of Bayesian search models," *Advances in Applied Probability* **25** (1993), pp. 714–716.
 22. A. J. Gross and M. Kamins, "Reliability assessment in the presence of reliability growth," *Annals of Assurance Sciences: 1968 Symposium on Reliability* (1968), pp. 406–416.
 23. T. Jayachadran and L. R. Moore, "A comparison of reliability growth models," *IEEE Transactions on Reliability* **R-25** (1976), pp. 49–51.
 24. N. L. Johnson, S. Katz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, Second Edition (John Wiley & Sons, Inc., New York, 1994).
 25. N. J. Lynn and N. D. Singpurwalla, "Comment: 'Burn-In' makes us feel good," *Statistical Science* **12** (1997), pp. 13–19.
 26. G. Manor and M. Kress, "Optimality of the greedy shooting strategy in the presence of incomplete damage information," *Naval Research Logistics* **44** (1997), pp. 613–622.
 27. T. A. Mazzuchi and R. Soyer, "A Bayes method for assessing product-reliability during development testing," *IEEE Transactions on Reliability* **42** (1993), pp. 503–510.
 28. National Research Council, *Reliability Issues for DoD Systems: Report of a Workshop*, Committee on National Statistics, Frank Samaniego and Michael Cohen, ed., The National Academies Press. Washington, D.C. (2002).
 29. D. E. Olsen, "Estimating reliability growth," *IEEE Transactions on Reliability* **R-26** (1977), pp. 50–53.
 30. S. M. Pollock, "A Bayesian reliability growth model," *IEEE Transactions on Reliability* **R-17** (1968), pp. 187–198.

31. E. A. Seglie, "Reliability growth programs from the operational test and evaluation perspective," *Reliability Growth Modeling: Objectives, Expectations, and Approaches*, ed. by K.J. Farquar and A. Mosleh. The Center for Reliability Engineering, Univ. of Maryland, College Park, MD 20742, (1998).
32. A. Sen, S. Basu and M. Banerjee, "Statistical analysis of life-data with masked cause-of-failure," *Handbook of Statistics: Reliability* (C. R. Rao and N. Balakrishnan, eds.) (2000, to appear).
33. W. M. Woods, "The effect of discounting failures and weighting data on the accuracy of some reliability growth models," *Proceedings of the Annual Reliability and Maintainability Symposium* (Los Angeles, CA, Institute of Electrical and Electronics Engineers, Inc.), (1990), pp. 200–204.
34. M. C. Yang, D. D. Wackerly and A. Rosalsky, "Optimal stopping rules in proof-reading," *J. Appl. Prob.* **19** (1982), pp. 723–729.

About the Authors

Professor Donald Gaver is a Distinguished Professor of Operations Research at the Naval Postgraduate School. He has a Ph.D. in Mathematics from Princeton University. He is a fellow of The American Statistical Association, of the Institute of Mathematical Statistics, and of the American Association for the Advancement of Science. He has served on several National Research Council panels. His research interests include stochastic and dynamic modeling, applications of statistics, and the testing and operational analysis of complex systems.

Professor Kevin Glazebrook holds the Chair of Management Science at Edinburgh University in Scotland. He has a Ph.D. from Cambridge University, England. His research interests range widely over Operations Research, Applied Probability and Statistics with a strong focus on sequential decision-making and stochastic optimization.

Professor Patricia Jacobs is a Professor of Operations Research at the Naval Postgraduate School. Her Ph.D. is in Applied Mathematics from Northwestern University. Her research interests include stochastic modeling and the operational analysis and testing of complex systems, particularly those in the military.

Dr. Ernest Seglie is the Science Advisor to the Director, Operational Test and Evaluation in the United States Department of Defense. He holds a Ph.D. in Physics from the University of Massachusetts at Amherst. He is the recipient of two meritorious civilian service awards from the United States Secretary of Defense.