NPS-OR-00-006

NAVAL POSTGRADUATE SCHOOL
Monterey, California

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

by

Donald P. Gaver
Kevin D. Glazebrook
Patricia A. Jacobs
Ernest A. Seglie

September 2000

Prepared for: Director, Operational Test and Evaluation (DOT&E)
Washington, DC 20301-1700

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943-5000

Rear Admiral D. R. Ellison
Superintendent

Richard Elster
Provost

This report was prepared for and funded by the Director, Operational Test and Evaluation (DOT&E), The Pentagon (Room 3E318), Washington, DC 20301-1700.

This report was prepared by:

DONALD P. GAVER
Professor of Operations Research

PATRICIA A. JACOBS
Professor of Operations Research

KEVIN D. GLAZEBROOK
Professor of Statistics

ERNEST A. SEGLIE
Science Director, DOT&E

Reviewed by:

R. KEVIN WOOD
Associate Chairman for Research
Department of Operations Research

RICHARD E. ROSENTHAL
Chairman
Department of Operations Research

Released by:

DAVID W. NETZER
Associate Provost and Dean of Research

REPORT DOCUMENTATION PAGE

Form approved

OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2000	Technical

4. TITLE AND SUBTITLE
Probability Models for Sequential-Stage System Reliability Growth *via* Failure Mode Removal

5. FUNDING
MIPR NO. DVAM00004

6. AUTHOR(S)
Donald P. Gaver, Patricia A. Jacobs, Kevin D. Glazebrook, Ernest A. Seglie

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-OR-00-06

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Director, Operational Test and Evaluation (DOT&E)
The Pentagon (Room 3E318)
Washington, DC 20301-1700

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Many systems, composed of hardware, software, and combinations thereof, function in sequential stages: each subsystem (stage) must operate correctly in order for the next to be challenged. All stages, including the interfaces between major function subsystems, are subject to design defects, which if actuated cause that stage, and hence that test, to fail. We provide models that evaluate the "testing as learning and improving" paradigm: the models describe the effect of end-to-end or linked-stage testing, and defect identification and removal, on field or delivered-system reliability. A major concern is the evaluation of operating characteristics of such test designs as the "first run of r total system successes (e.g. 3)" stopping rule. The models include Bayesian formulations in which the unknown number of defects in each subsystem at any stage during testing is a random variable with known distribution. The models and methods of this paper provide test planners with the answers to "what if" questions concerning the likely future(s) of entire systems placed on test. They can be used to address test resource requirements.

14. SUBJECT TERMS	15. NUMBER OF PAGES
operational test and evaluation; reliability growth; system with substages; software testing	57
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

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

Donald P. Gaver
dgaver@nps.navy.mil
Operations Research Department
Naval Postgraduate School
Monterey, California

Patricia A. Jacobs
pajacobs@nps.navy.mil
Operations Research Department
Naval Postgraduate School
Monterey, California

Kevin D. Glazebrook
kevin.glazebrook@ncl.ac.uk
Department of Statistics
Newcastle University
Newcastle upon Tyne, UK

Ernest A. Seglie
eseglie@dote.osd.mil
The Pentagon
Washington, DC

Abstract

Many systems, composed of hardware, software, and combinations thereof, function in sequential stages: each subsystem (stage) must operate correctly in order for the next to be challenged. All stages, including the interfaces between major function subsystems, are subject to design defects, which if actuated cause that stage, and hence that test, to fail. We provide models that evaluate the "testing as learning and improving" paradigm: the models describe the effect of end-to-end or linked-stage testing, and defect identification and removal, on field or delivered-system reliability. A major concern is the evaluation of operating characteristics of such test designs as the "first run of r total system successes (e.g. 3)" stopping rule. The models include Bayesian formulations in which the unknown number of defects in each subsystem at any stage during testing is a random variable with known distribution. The models and methods of this paper provide test planners with the answers to "what if" questions concerning the likely future(s) of entire systems placed on test. They can be used to address test resource requirements.

1. Introduction and Model Formulation

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 Seglie, 1998). 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, $\mathbb{S}$, 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; $s = 1, 2, \ldots, S$. If a demand at an intermediate stage/subsystem, s, succeeds, i.e. any faults 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 faults activate). That is, a system following current design and realization must function "end-to-end" in order to operate reliably — "suitably" in military parlance. This "success" does *not* mean that the particular system mission is necessarily overall successful or "effective" (a reliable weapon may not accomplish its mission: it may miss, or hit a wrong target). Such may also be, in part, reliability issues, but attributable to

C4ISR errors. Note, however, that the design defect removals we aim for may include those in basic functionality ("effectiveness") such as accuracy and lethality.

To perform a system-level operational test of $\mathbb{S}$, 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 Bogdonavicius and Nikulin (2000). 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 $\overline{\theta}_s$ if a demand is received at that stage; otherwise are inactive or survive with probability $1 - \overline{\theta}_s = \theta_s$. In order for the s^{th} subsystem to *experience test*, and hence possibly reveal a failure mode, all previous $i \in (1, 2, \ldots, s-1)$ 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, but this simplicity may not hold: new failure modes may actually be introduced, and bedrock non-removable failure modes will remain. These realities are ignored for simplicity in the present report, so the results are likely to be optimistic. We also ignore the detrimental effects of system aging (one-shot items eventually age on the shelf). We again emphasize that in operational field testing it is often the inter-subsystem *interactions* that exhibit surprising new failure modes which must be discovered by suitable testing. Our model can cover such situations by simply defining some stages as "interaction subsystems".

Here, testing the complete system (e.g. a missile or an information system) requires that "early" ($s = 1, 2, \ldots$) subsystems survive so that "late" ($s = \ldots S\text{-}1, S$) subsystems can experience demand, and hence literally be subjected to test. Failures of early subsystems

protect later subsystems from test; this effect must be overcome in order for the entire system to be tested. Engineering-level or developmental tests (DT) of the individual component subsystems will be, or have been, carried out, but these cannot be completely trusted to identify all failure modes that may appear in actual operation when the entire system is assembled and tested, much less in the field. In 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 Block and Savits (1997) for a nice review, and also Lynn and Singpurwalla (1997). 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.

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 starting expensive field testing:

(a) After a given number of *system* tests, what is the (approximate) probability that the system will operate satisfactorily (not fail) 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 (e.g. 3, or 5) *consecutive* end-to-end test successes, or, in statistical parlance, a (first) *run of r*; 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 (possibly fixed, or random governed by a stopping rule such as "first occurrence of a run of r (e.g. $r = 3, 5, \ldots$ whatever) successful

reliability tests of entire system"), after which 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 Fries and Maillart (1996). What is the probability that the system completes a mission that requires at least M successes if $M+R$ systems are allocated? What is the mean, and variability, of the number of tests required?

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

A system is composed of a number, S, of subsystems each of which contains an uncertain number of failure modes (design defects). When a design defect is activated during a test, the system fails at that subsystem, and that particular test terminates. Figure 1 illustrates the configuration, and outcomes.

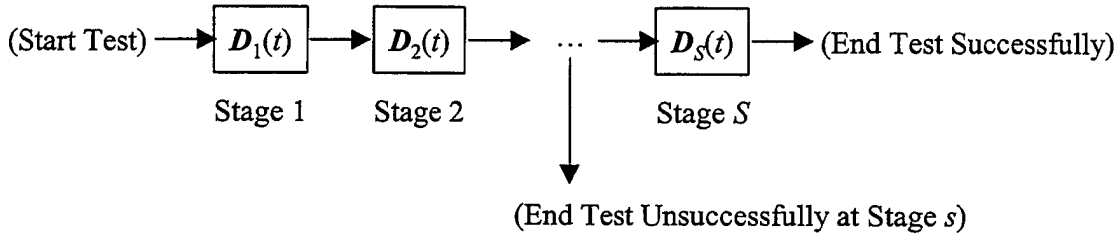

Figure 1

Let

$\boldsymbol{D}_s(t)$ = number of design defects present in Stage s after t tests;

$\boldsymbol{D}_s(0)$ = number of design defects present initially in Stage s before test (or at least before operational, end-to-end, testing). At this stage the distribution of $D_s(0)$ *may* be treated as a Bayes prior, *or* as an expression of inherent variability; a Bayes prior can describe hyperparameters.

Stages are *request-activated strictly serially*, starting with Stage 1 and ending with Stage S. If any stage fails to respond, the following stage is not demanded/request-activated and the system trial/test fails. The Stage s request activation can only occur if *all* previous stages, $i = 1, 2, \ldots, s-1$ respond to their request activations. (This does *not* imply that stages that successfully respond to their request activations are free of defects – they may well randomly activate and cause failures on later trials, or even on a mission after system release.) Repeated testing *tends* to remove defects, but there will be little reward from testing long enough to eliminate defects that are unlikely to activate in the field.

4. Model: Stage-wise (Binomial) Failures, One-at-a-Time Removable

Invoke the stage-wise failure model, and allow *only one activated design defect* to be identified and removed per test (no new design defects are introduced). If more than one failure mode or defect is activated on a test we assume that only one of these can be identified and removed; the others can activate again.

There follow several functional-equation modeling systems that respond to questions posed earlier.

4.1 Expected Probability of System Field Success After a Fixed Number, *t*, of Tests

Let $\boldsymbol{D}_i(0)$ be the initial number of defects in stage i, $i = 1, \ldots, S$. Assume that some defect in stage i is revealed during a test with probability $\tilde{p}_i(d_i)$; $\tilde{q}_i(d_i) = 1 - \tilde{p}_i(d_i)$ where d_i is the number of defects in stage i, assumed ≥ 1. A special case is $\tilde{q}_i(d_i) = \theta_i^{d_i}$, but allowance for random (extra) variability in θ (e.g. as a Beta random variable) is natural to reflect within-stage variability beyond the simple binomial. Note that this is viewed as representing physical mixtures; it is not necessarily a Bayesian prior. A defect revealed in stage i causes the system to fail without revealing any defects in later stages; these are *hidden* for that test. Each test reveals *at most one* identifiable defect or fault. Such a discovered defect is assumed removed *with certainty* (with probability equal to

one) by present assumption (if it is successfully removed with probability ρ then we may replace the probability a defect is discovered and removed, e.g. $1-\theta$ by $(1-\theta)\rho$ and proceed). Let $\boldsymbol{D}_i(t)$ be the number of defects remaining in stage i after t tests. Let Q_i be the probability a remaining defect in stage i does not activate while the system is put in use *in the field* during one mission. (Desirably, $Q_i \approx$ or $> \tilde{q}_i$, the probability a test does not reveal a defect.) Note: for *initial example*, but *not throughout*, activation of some design fault or defect is Binomial: $\tilde{p}_i(d_i) = 1-\theta_i^{d_i}$. It is possible to represent extra-Binomial stage-to-stage variability by mixing within stages to provide extra-binomial variability: $E\left[\theta_i^{d_i}\right]$; see Appendix A.

Define

$$p(d_1,\ldots d_S,t) = P\left\{\boldsymbol{D}_1(t)=d_1,\ldots,\boldsymbol{D}_S(t)=d_S\right\}, \tag{4.1}$$

the joint probability of the number of defects present in each stage after t tests. The following forward equation (Markov chain) can be established by conditioning:

$$p(d_1,\ldots,d_S,t+1) =$$

$$\underbrace{p(d_1,\ldots,d_S,t)\prod_{i=1}^{S}\tilde{q}_i(d_i)}_{\substack{\text{No design defects}\\ \text{removed on test } t}} + \underbrace{\sum_{i=1}^{S} p(d_1,\ldots,d_i+1,\ldots,d_S,t)\left(\prod_{j=1}^{i-1}\tilde{q}_j(d_j)\right)^{*}\left[1-\tilde{q}_i(d_i+1)\right]}_{\substack{\text{One design defect removed on test } t\\ \text{(e.g. from stage } j,\ j=1,2,\ldots,S)}} \tag{4.2}$$

Note: the term in the last product, ( )* = 1 if $i = 1$.

The recursion is initialized with

$$p(d_1,\ldots,d_S,0) = \begin{cases} 1 & \text{if } \boldsymbol{D}_1(0)=d_1,\ldots,\boldsymbol{D}_S(0)=d_S \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

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

$$\tilde{Q}(t) = \sum_{d_1,\ldots,d_S} p(d_1,d_2,\ldots,d_S,t)\prod_{j=1}^{S} Q_j^{d_j}\ . \tag{4.4}$$

Or, more generally, with field probabilities of survival allowed to differ from those of the test *and* experience stage-wise mixing,

$$\tilde{Q}(t) = \sum_{d_1,\ldots,d_S} p(d_1,d_2,\ldots,d_S,t)\prod_{j=1}^{S}\tilde{Q}_j(d_j). \tag{4.5}$$

The following examples are based on simple Binomial stagewise failures: $\tilde{q}_i(d_i) = q_i^{d_i}$, $\tilde{Q}_i(d_i) = q_i^{d_i}$.

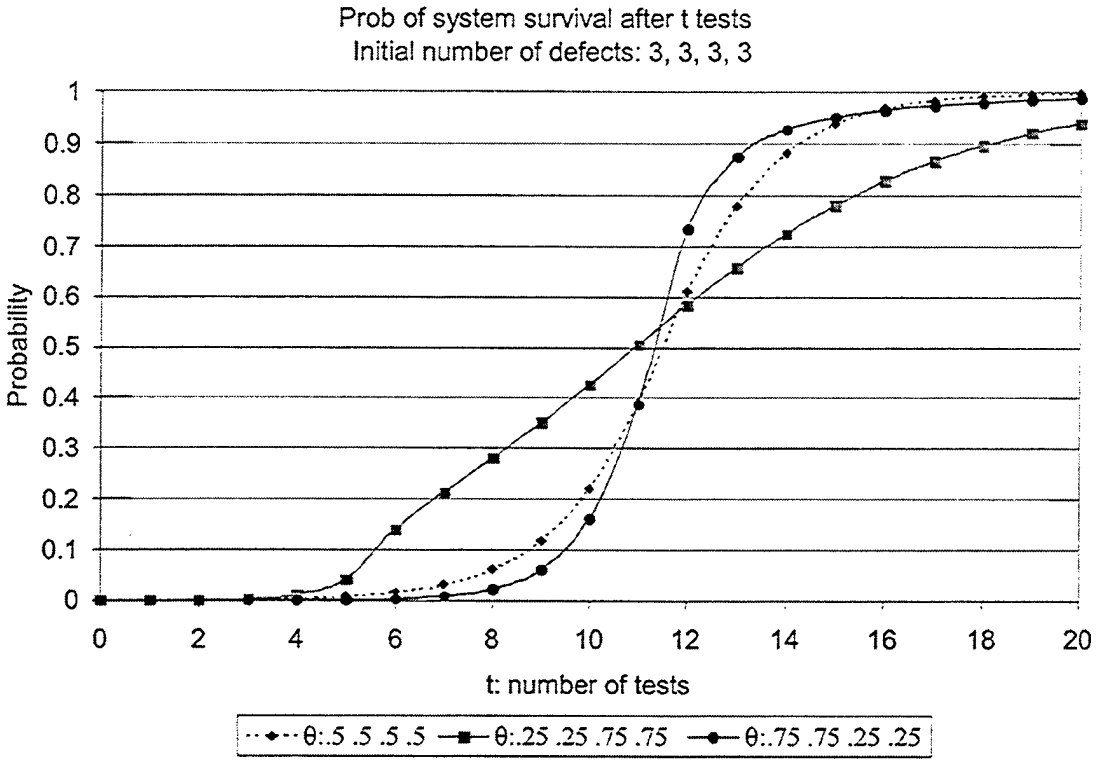

Figure 2

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. Fries (1993). 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 need to re-test the last stages forces more tests of the earlier

because of the end-to-end success requirement. It is unlikely that a designer, or tester, can ever directly influence such a distribution of defects, but there may be implications for variations in the intensity of component-level testing: one might tolerate a few more faults in later stages, so that earlier stages will be subjected to more operational end-to-end tests.

4.2 Probability of Mission Success in the Field if Testing Stops After First Run of *r* Consecutive Successful Tests

Suppose the system test is stopped when there are r ($r \geq 1$, e.g. 3) 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. The probability of system survival after completion of the test can be computed using a backward equation as follows. Let $p_r(d_1, \ldots, d_S)$ be the conditional probability of system mission survival in the field after the test, given that the initial numbers of defects are $\boldsymbol{D}_1(0) = d_1, \ldots, \boldsymbol{D}_S(0) = d_S$. Use the previous stagewise survival probabilities, $\tilde{q}_i(d_i)$ to write

$$p_r(d_1,d_2,\ldots,d_S) = \underbrace{\left(\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)^r \prod_{i=1}^{S}\tilde{Q}_i(d_i)}_{\substack{\text{Run of } r \text{ successful tests} \\ \text{occurs before any test failures}}} \tag{4.6}$$

$$+\underbrace{\left[1-\left(\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)^r\right]}_{\text{No } r\text{-run during first } r \text{ tests}} \underbrace{\sum_{i=1}^{S}\left(\prod_{j=1}^{i-1}\tilde{q}_j(d_j)\right)^*\left(1-\tilde{q}_i(d_i)\right)\left[1-\left(\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)\right]^{-1} p_r(d_1,\ldots,d_i-1,\ldots,d_S)}_{\text{Start over after a failure at stage } i \text{ before run of } r \text{ successful tests achieved}}$$

Note: $(\quad)^* = 1$ if $i = 1$

The recursion starts with $p_r(0, \ldots, 0) = 1$, and thus builds up to any desired initial load of design defects.

$$\tilde{q}_i(d_i) = E\left[\theta^{d_i}\right]$$

Here the field survival probability is assumed equal to the field test system survival probability: $\tilde{Q}(d) = \tilde{q}(d)$.

Note: the above equation permits quick numerical determination of the *mean* or unconditional probability of field success. Simulations show that there can be substantial difference between the mean and the actual probability of success, depending on fault survival. The following forward equation can be used to calculate the distribution of the probability of system survival after the test.

Let $\tilde{\gamma}_r(a_1,\ldots,a_S)$ be the probability that there are a_i, $i = 1, \ldots, S$ defects remaining sometime during the test. The probabilities $\tilde{\gamma}_r(a_1,\ldots,a_S)$ can be obtained recursively as follows.

$$\tilde{\gamma}_r(a_1,\ldots,a_S)$$

$$= \sum_{s=1}^{S} \tilde{\gamma}_r(a_1,\ldots,a_s+1,\ldots,a_S) \frac{\left[1-\left(\left(\prod_{\substack{i=1\\ i\neq s}}^{S} q_i(a_i)\right) q_s(a_s+1)\right)^r\right]}{\left[1-\left(\left(\prod_{\substack{i=1\\ i\neq s}}^{S} q_i(a_i)\right) q_s(a_s+1)\right)\right]} \left(\prod_{i=1}^{s-1} q_i(a_i)\right)^{*} \left[1-q_s(a_s+1)\right] \tag{4.7}$$

with initial condition $\tilde{\gamma}_r(d_1,\ldots,d_S) = 1$ where d_i is the initial number of defects in stage i, $i = 1, \ldots, S$; note $(\ )^{*} = 1$ if $s = 1$.

The probability of having a_i remaining defects in stage i, $i = 1, \ldots, S$ after completion of the test is

$$\gamma_r(a_1,\ldots,a_S) = \tilde{\gamma}_r(a_1,\ldots,a_S)\left[\prod_{i=1}^{S} q_i(a_i)\right]^r \tag{4.8}$$

These probabilities can be used to obtain the distribution of the probability of field survival after the test is completed.

Numerical results with Bernoulli-trials Stagewise Variability.

In the example whose results are displayed in Figure 3 $\tilde{q}_i(d_i) = \tilde{Q}_i(d_i) = \theta_i^{d_i}$. Note that Figure 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 $\theta_1 = 0.75$, $\theta_2 = 0.25$, $\theta_3 = 0.75$, $\theta_4 = 0.25$, the probability of mission survival after testing increases slightly as the initial number 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.

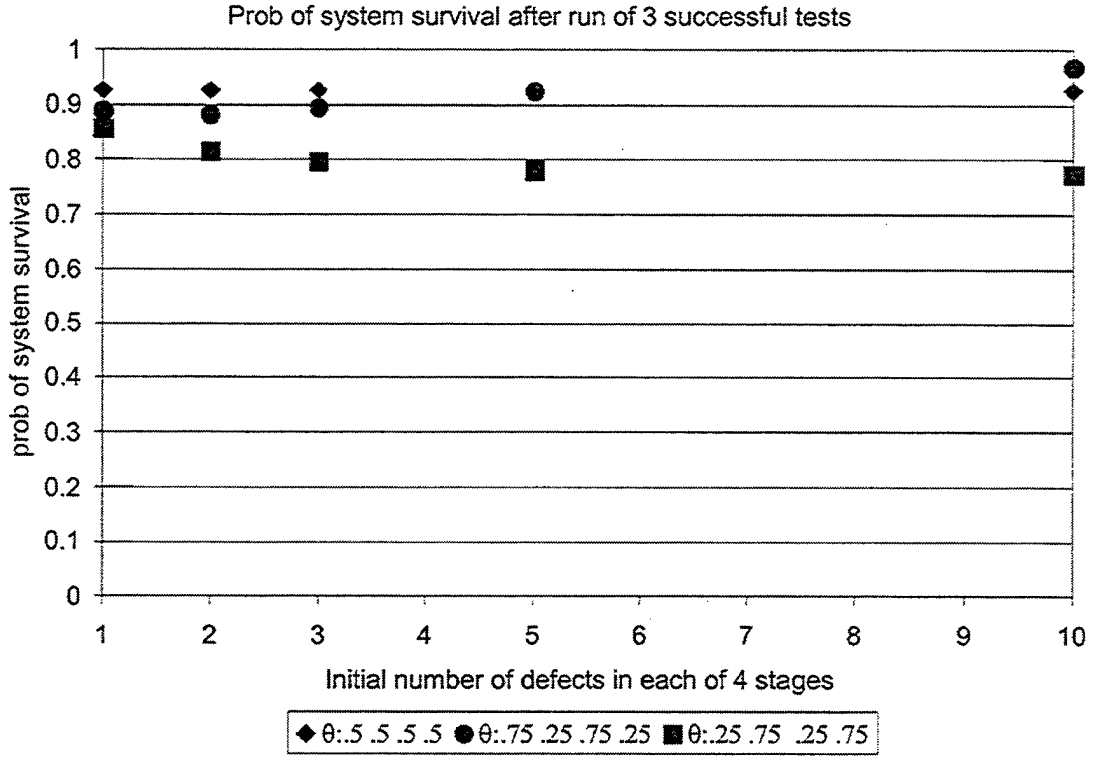

Figure 3

4.3 Mean Time to Stop After Reaching First Run of r Consecutive Successful Tests

This is a measure of the time cost of a test program that is run-terminated. Let $r_r(d_1, d_2, \ldots, d_S)$ be the conditional expected time (number of tests) until a run of r successes first occurs, given that there are initially d_i defects in stage i. Here is a backward equation for this performance measure.

$$r_r(d_1,d_2,\ldots,d_S) = \underbrace{r\left(\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)^r}_{\substack{r\text{-run after } r \text{ tests:}\\ \text{initial } r\text{-run}}}$$

$$+\underbrace{\left(1-\left(\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)^r\right)}_{\text{Probability no initial run of } r}\times\underbrace{\sum_{n=1}^{r}\left\{n+\sum_{j=1}^{S}r_r(d_1,\ldots,d_j-1,\ldots,d_S)\frac{\left(\prod_{i=1}^{j-1}\tilde{q}_i(d_i)\right)^*\left(1-\tilde{q}_j(d_i)\right)}{1-\prod_{i=1}^{S}\tilde{q}_i(d_i)}\right\}}_{\text{Expected number of tests to achieve } r\text{-run, given failure to achieve initial } r\text{-run}}$$

$$\times\underbrace{\frac{\left(\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)^{n-1}\left(1-\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)}{1-\left(\prod_{i=1}^{S}\tilde{q}_i(d_i)\right)^r}}_{\substack{\text{Probability first activation after } n\le r \text{ tests,}\\ \text{given no run of } r}} \tag{4.9}$$

Note: $(\quad)^* = 1$ if $j = 0$

An initial condition is

$$r_r(0, 0, \ldots, 0) = r \tag{4.10}$$

Numerical results with stagewise over-variability.

In the examples of Figures 4, 6, and 8 we compare the probability of mission success for several cases when (a) the stagewise probabilities θ_i are taken as invariable ("fixed"), Bernoulli trials probabilities vs. (b) the stagewise probabilities are themselves variable, independently for each stage and test. This extra-Binomial variability can conveniently be characterized by a diffuse beta distribution with mean equal to the fixed values of (a). The

variability in (b) represents *some aspects of* uncontrollable *between-test*, and *between-test-stage* conditions; see Appendix A. In each case that has been investigated the probability of field success is higher for (a), "fixed" or controlled probabilities, than for (b), the corresponding stage-wise mixed probability. Practical considerations suggest that (b) may be the more qualitatively realistic, because of the likelihood of extra uncontrollable variations in the field. Some such are likely to be roughly common to all stages; this is analyzed in Appendix B.

From Figures 4, 6, and 8 it is striking that the *order* of the defect survival probability occurrence (which may be practically difficult to control at the developmental testing stage) can be influential at the final field survival probability level. Once again, the case of Figure 8, θ: 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. From Figures 5, 7, and 9 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.

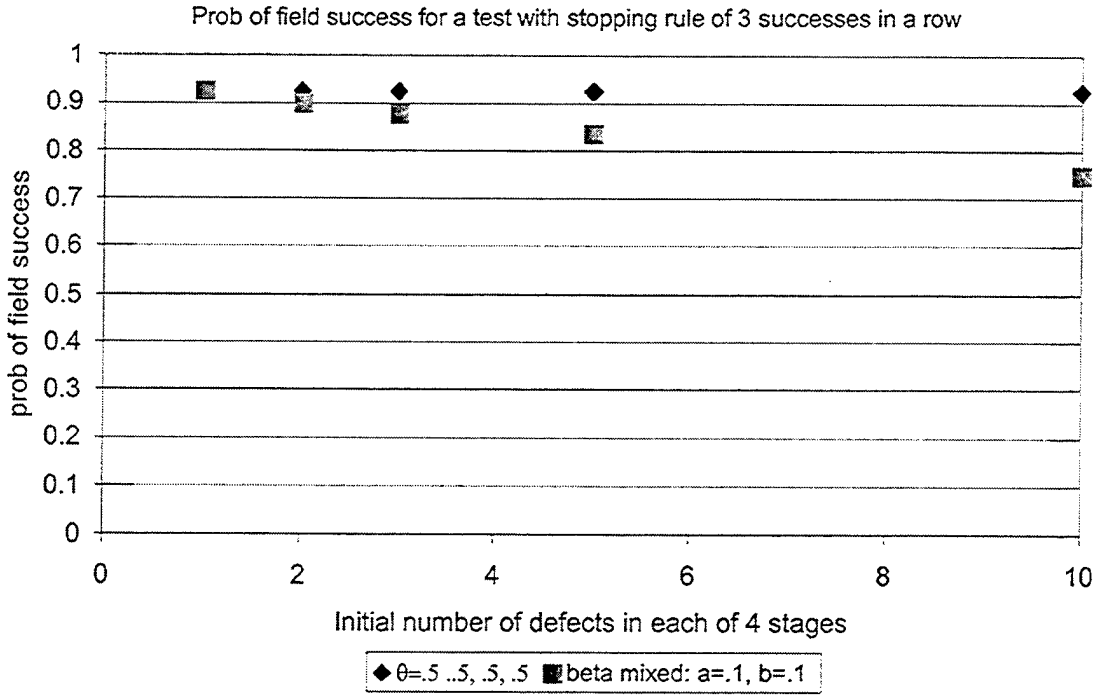

Figure 4

The mixing distribution employed in Figure 4 is symmetric, but with high weighting near 0 and 1. Such an environment badly penalizes the tester if there are many (e.g. 5 or more) defects in the system initially.

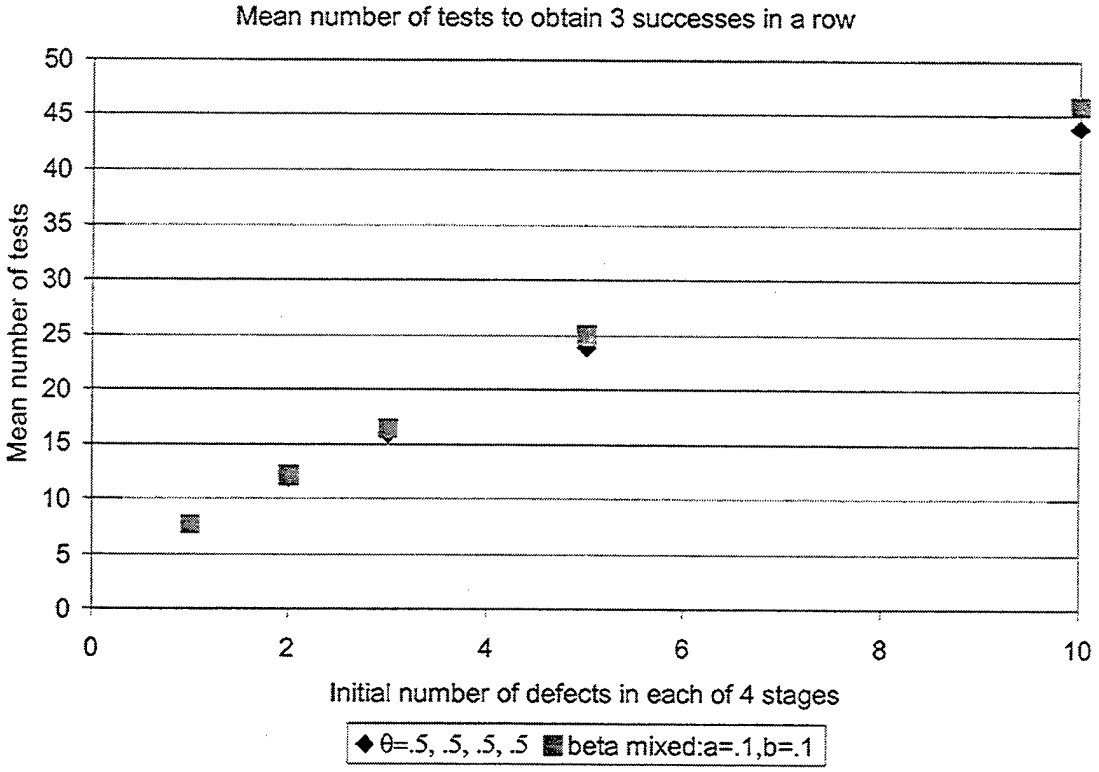

Figure 5

The *expected* times to complete the tests in Figure 5 are remarkably similar for these cases.

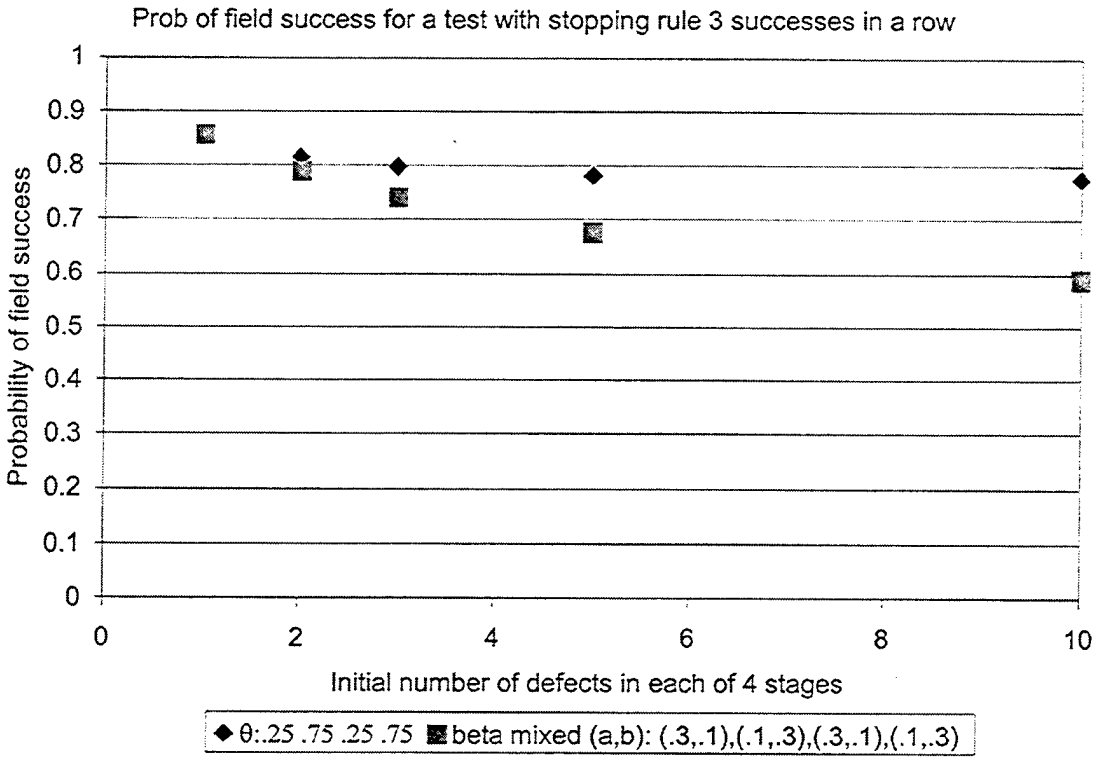

Figure 6

In Figure 6 a somewhat diffuse mixing distribution (Beta) is used for each stage, with means located at the "deterministic" levels. Once again, however, the stagewise mixtures at stages, independent, and recalculated independently between tests, have a substantial degrading effect on the mean probability of a system's field success if the system is accepted after a run of 3.

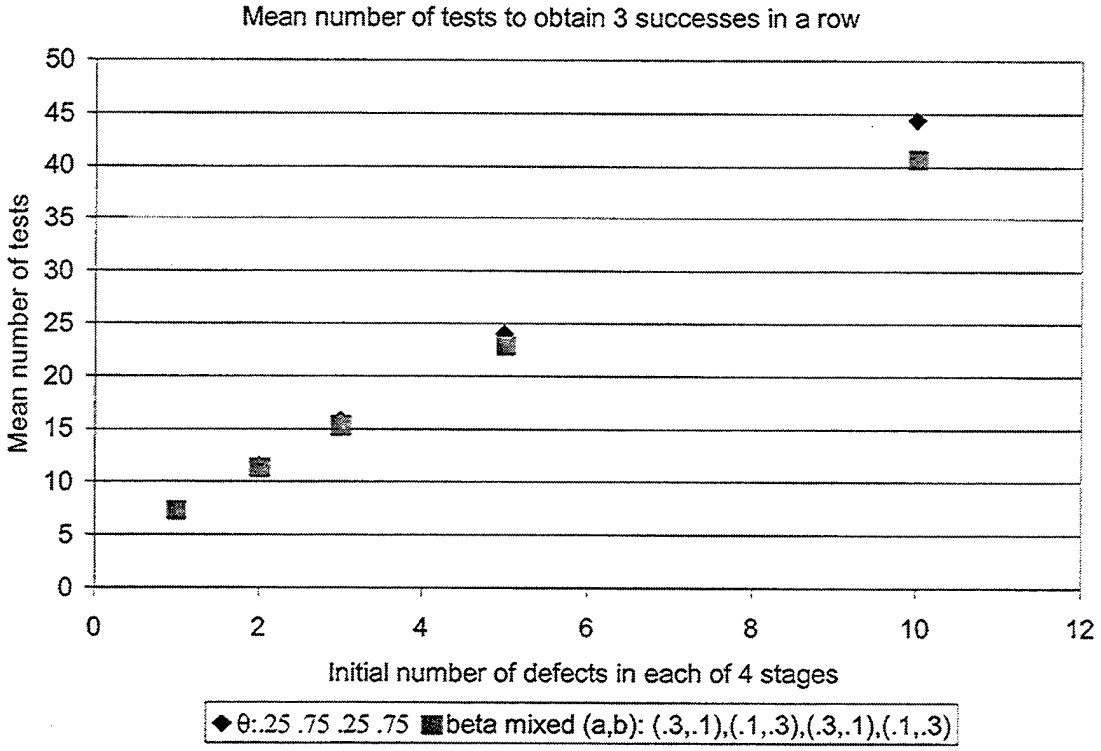

Figure 7

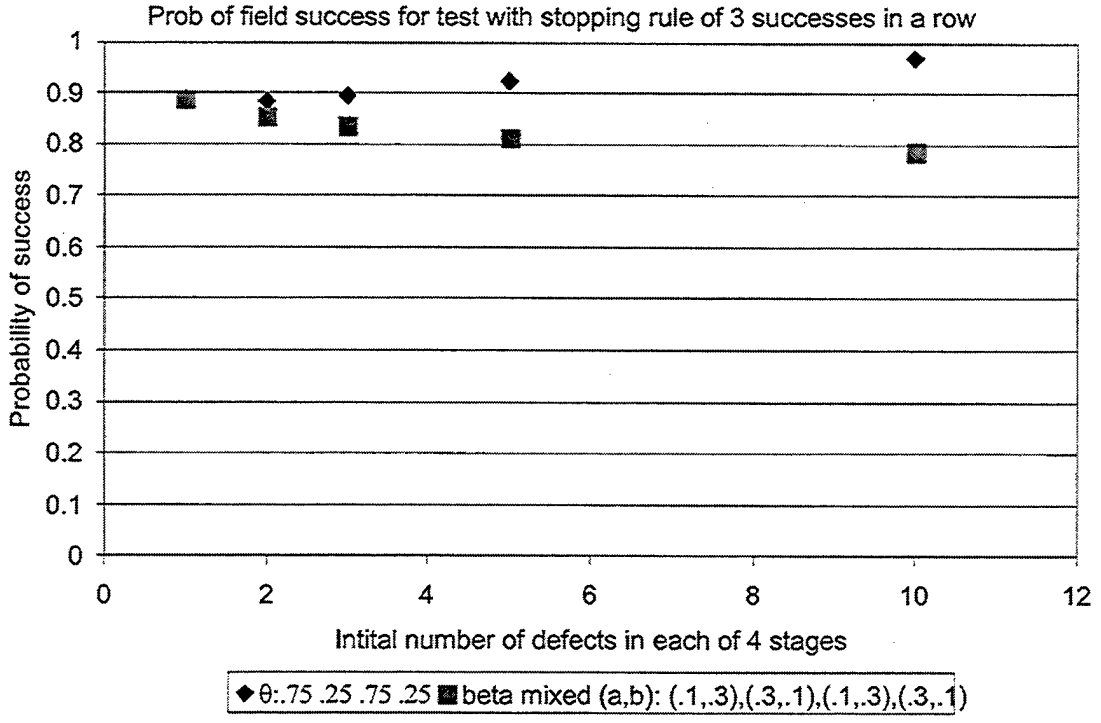

Figure 8

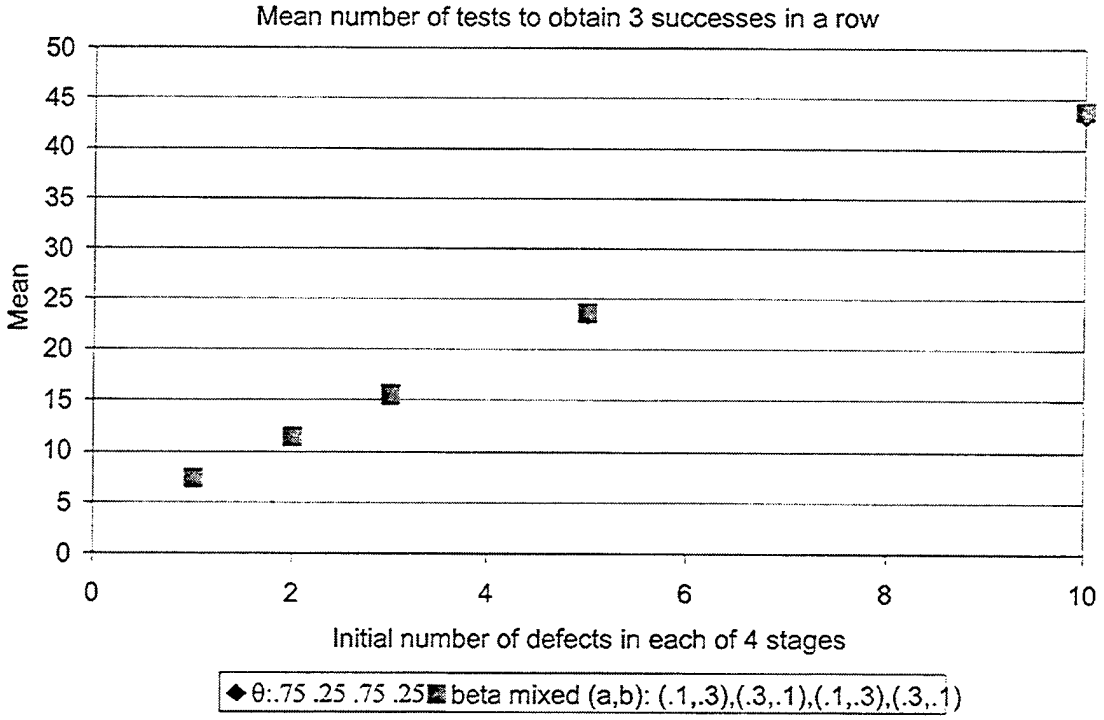

Figure 9

5. Bayesian Formulations

A natural approach to the uncertainty concerning the numbers of design defects initially present is a Bayesian one in which $D_i(0)$ is treated as a random variable with (prior) distribution $\Pi^i = \left\{\Pi^i_d, d \geq 0\right\}$, $1 \leq i \leq S$. In what follows we shall suppose that the random variables $D_i(0)$, $1 \leq i \leq S$ are independent and that the conditional model for failure discovery and removal is as in Section 4 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 $H_{it} = \{x_{i1}, x_{i2}, \ldots, x_{it}\}$ 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 $D_i(t)$ upon suitable application of Bayes' theorem. Updating is described as follows:

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

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.

In this highly complex scenario it seems reasonable to make an initial search for simplicity. In particular, we seek conditions under which each $\Pi^i(t)$ depends upon H_{it} *only* through $\{S_i(t), F_i(t)\}$, where

$$S_i(t) = \sum_{j=1}^{t} 1(x_{ij} = S_{ij}),$$

the total number of successful operations of subsystem i during t tests and $F_i(t)$ is similarly defined in terms of failures (i.e. defect activations and removals). Expressed simply, we require that the numbers of successes and failures to date should be sufficient statistics for each subsystem. Work by Benkherouf and Bather (1988) in the context of oil exploration implies that this requirement forces a conditional model of the form

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

which is the Binomial case of Section (4.1). Until indicated otherwise we suppose that (5.2) holds.

Further, Glazebrook (1993) introduced a family of prior distributions which are conjugate for this problem. Let $D_i(0) \sim \Pi^i(0) \equiv \Pi(\lambda_i, \theta_i, \phi_i)$, $1 \leq i \leq S$, where the probability mass function (p.m.f.) for $\Pi(\lambda, \theta, \phi)$ is given by

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

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. Parameter θ is always assigned the value of the probability in the conditional Binomial model in (5.2). Special cases of this model are

$$\Pi(\lambda, \theta, 0) \equiv E(\lambda, \theta); \quad \Pi(\lambda, \theta, 1) \equiv H(\lambda, \theta), \tag{5.4}$$

the Euler and Heine distributions with parameters (λ, θ) respectively. These are discussed by Benkherouf, Glazebrook, and Owen (1992). The reader should note that, for regions of its parameter space, the Euler distribution may approach either a Poisson or a geometric distribution. Thus the prior family (5.3) is quite versatile.

With the prior $\Pi(\lambda_i, \theta_i, \phi_i)$ in (5.3) and the conditional model (5.2), the posterior distribution for $D_i(t)$ is given (upon operation of (5.1)) by

$$\Pi^i(t) \equiv \Pi\left\{\lambda_i \theta_i^{S_i(t)+\phi_i F_i(t)}, \theta_i, \phi_i\right\}, \quad 1 \leq i \leq S. \tag{5.5}$$

From (5.5), the situation following t tests is such that the overall rates of defect detection in subsystems have fallen to the values

$$\lambda_i \theta_i^{S_i(t)+\phi_i F_i(t)}, \quad 1 \leq i \leq S; \tag{5.6}$$

it is reasonable to stop testing when these values are sufficiently small.

5.1 Bayes stopping rules

With the above structures in place we can design stopping rules which are Bayes optimal for a range of objectives. In such problems, a major difficulty is presented by the computational demands of producing a fully sequential solution *via* dynamic programming. However, in the context of reliability growth in which the number of defects is reduced (stochastically) at each test, one-step-look-ahead (myopic) rules should perform well. See, for example, Benkherouf and Bather (1988), Gittins (1989) and Manor and Kress (1997) for examples of this phenomenon. Suppose that we wish to maximize an objective

$$E_{\Pi}\left\{-cT+\tilde{Q}_T\right\} \tag{5.7}$$

where T is the number of tests performed, c is a (suitably standardized) cost per test and

$$\tilde{Q}_T=\begin{cases}1, & \text{for successful field deployment following } T \text{ tests,}\\ 0, & \text{otherwise.}\end{cases}$$

In (5.7) the expectation is taken with respect to the prior distributions (summarized by Π).

A key quantity required for the development of a solution to (5.7) is the *predictive* probability of successful field deployment $Q(\mathbf{S}, \mathbf{F})$ at a point in which the data from testing are summarized by sufficient statistics $\mathbf{S}, \mathbf{F} = \{(S_i, F_i),\ 1 \le i \le S\}$. Utilizing the above independence assumptions,

$$Q(\mathbf{S},\mathbf{F})=\prod_{i=1}^{S} Q_i(S_i,F_i) \tag{5.8}$$

where $Q_i(S_i, F_i)$ is the predictive probability of successful field deployment of subsystem i with sufficient statistics (S_i, F_i), $1 \le i \le S$. Taking $\Pi^i(0) \equiv \Pi(\lambda_i, \theta_i, \phi_i)$, conditional model (5.2) and $\tilde{Q}_i(d)=\theta_i^d$, $d \ge 0$, $1 \le i \le S$, we deduce from (5.5) that

$$Q_i(S_i,F_i)=\sum_{d\geq 0}\left(1-\theta_i^d\right)\Pi_d\left\{\lambda_i\theta_i^{S_i+\phi_i F_i},\theta_i,\phi_i\right\}$$

$$=1-\lambda_i\theta_i^{S_i+\phi_i F_i}\,\frac{\Pi_0\left(\lambda_i\theta_i^{S_i+\phi_i F_i},\theta_i,\phi_i\right)}{\Pi_0\left(\lambda_i\theta_i^{S_i+\phi_i(F_i+1)},\theta_i,\phi_i\right)}. \tag{5.9}$$

We write $E\{Q(\mathbf{S},\mathbf{F})\}$ for the expected predictive probability of successful field deployment following *one additional test*, starting from $(\mathbf{S},\mathbf{F})$. Utilizing (5.5) and (5.8) we deduce that

$$E\{Q(\mathbf{S},\mathbf{F})\}=\sum_{i=1}^{S}\left\{\prod_{j=1}^{i-1}Q_j(S_j,F_j)\right\}\{1-Q_i(S_i,F_i)\}\,Q\left(\mathbf{S}+\sum_{j=1}^{i-1}\mathbf{1}^j,\mathbf{F}+\mathbf{1}^i\right)$$
$$+\left\{\prod_{i=1}^{S}Q_i(S_i,F_i)\right\}Q\left(\mathbf{S}+\sum_{i=1}^{S}\mathbf{1}^i,\mathbf{F}\right) \tag{5.10}$$

where in (5.10), $\mathbf{1}^i$ is an S-vector whose i^{th} component is one, with zeros elsewhere. A Bayes myopic stopping rule for problem (5.7) concludes testing as soon as the gain in system reliability from one further test is less than the cost of the test. Formally, the associated stopping region is

$$[(\mathbf{S},\mathbf{F});\,E\{Q(\mathbf{S},\mathbf{F})\}-Q(\mathbf{S},\mathbf{F})\,\}<c]. \tag{5.11}$$

When c is small (i.e. the cost of one more test is negligible compared to the utility of having a system fully operational in the field) then from (5.9) and (5.10) we can show that the stopping region in (5.11) may be well approximated by

$$\left[(\mathbf{S},\mathbf{F});\left\{\sum_{i=1}^{S}\lambda_i\theta_i^{S_i+\phi_i F_i}(1-\theta_i)\right\}Q(\mathbf{S},\mathbf{F})<c\right] \tag{5.12}$$
$$\supseteq\left\{(\mathbf{S},\mathbf{F});\sum_{i=1}^{S}\lambda_i\theta_i^{S_i+\phi_i F_i}(1-\theta_i)<c\right\}.$$

See the comments around (5.6) above. The simple conservative stopping rules above will approximate (5.11) well for c close to zero.

Numerical example

The stopping rule in (5.12) was applied to a testing problem with 4 stages, 3 defects being initially present in each stage. Results may be found in Table 1. Four different sets of theta values were considered, namely, $\theta(1)$: 0.75, 0.75, 0.75, 0.75, $\theta(2)$: 0.5, 0.5, 0.5, 0.5, $\theta(3)$: 0.25, 0.25, 0.75, 0.75, and $\theta(4)$: 0.75, 0.75, 0.25, 0.25. The prior distributions used to determine the stopping rules were taken to be Euler in all cases. For these distributions a value of λ is required. We explored two different approaches to making this choice. In approach 1 we set λ to be $1 - \theta^3$, thus guaranteeing that the unconditional initial subsystem failure probability in the Bayesian model was equal to the (actual) initial failure probability with three defectives. Under this approach, we used priors $E(0.578, 0.75)$, $E(0.875, 0.5)$, and $E(0.984, 0.25)$ as appropriate. These λ choices are denoted $\lambda(\bullet 1)$ in the table. Under approach 2, for given θ we chose λ such that the mean of the $E(\lambda, \theta)$ distribution was 3. Under this approach, we used priors $E(0.500, 0.75)$, $E(0.672, 0.5)$, and $E(0.734, 0.25)$ with the λ choices denoted $\lambda(\bullet 2)$ in the table. Testing continued until an appropriate version of the stopping criterion in (5.12) was met with c set equal to 0.1, 0.06, 0.04, 0.02, 0.01. The smaller the value of c, the more conservative is the resulting test. Each case (cell of the table) was run 3,000 times. The results are given by the unbracketed values in each cell of the table, which are (reading from top to bottom):

(1) actual mean probability of field success at end of testing;

(2) mean predictive probability of field success under the Bayesian model at end of testing;

(3) mean number of tests.

The bracketed values are the corresponding standard errors.

Table 1
Probability of field success and mean number of tests with a Bayes myopic stopping rule

Scenario \ c	0.1	0.06	0.04	0.02	0.01
	0.515 (0.004)	0.736 (0.004)	0.842 (0.003)	0.929 (0.002)	0.966 (0.002)
$(\theta(1),\lambda(11))$	0.475 (0.001)	0.712 (0.000)	0.831 (0.000)	0.927 (0.000)	0.960 (0.000)
	11.569 (0.025)	15.041 (0.014)	17.415 (0.011)	20.712 (0.009)	22.970 (0.003)
	0.437 (0.003)	0.724 (0.004)	0.829 (0.003)	0.925 (0.002)	0.961 (0.002)
$(\theta(1),\lambda(12))$	0.467 (0.001)	0.737 (0.000)	0.848 (0.000)	0.921 (0.000)	0.965 (0.000)
	10.562 (0.024)	14.831 (0.018)	17.263 (0.015)	19.906 (0.006)	22.847 (0.007)
	0.865 (0.004)	0.928 (0.003)	0.938 (0.003)	0.969 (0.002)	0.987 (0.001)
$(\theta(2),\lambda(21))$	0.777 (0.001)	0.882 (0.000)	0.937 (0.000)	0.969 (0.000)	0.984 (0.000)
	13.904 (0.006)	14.958 (0.004)	15.872 (0.007)	16.937 (0.004)	17.973 (0.003)
	0.750 (0.005)	0.867 (0.004)	0.933 (0.003)	0.968 (0.002)	0.983 (0.002)
$(\theta(2),\lambda(22))$	0.801 (0.000)	0.902 (0.000)	0.952 (0.000)	0.976 (0.000)	0.988 (0.000)
	13.426 (0.014)	14.712 (0.010)	15.862 (0.007)	16.935 (0.005)	17.966 (0.003)
	0.270 (0.002)	0.608 (0.005)	0.843 (0.003)	0.931 (0.002)	0.968 (0.002)
$(\theta(3),\lambda(31))$	0.251 (0.002)	0.564 (0.004)	0.823 (0.000)	0.921 (0.000)	0.965 (0.000)
	7.952 (0.018)	12.609 (0.053)	16.677 (0.010)	19.886 (0.006)	22.871 (0.006)
	0.253 (0.002)	0.691 (0.004)	0.815 (0.003)	0.918 (0.002)	0.963 (0.002)
$(\theta(3),\lambda(32))$	0.313 (0.002)	0.710 (0.000)	0.832 (0.000)	0.929 (0.000)	0.961 (0.000)
	7.761 (0.014)	13.738 (0.017)	16.262 (0.014)	19.673 (0.010)	21.937 (0.005)
	0.860 (0.004)	0.908 (0.003)	0.923 (0.003)	0.947 (0.002)	0.964 (0.002)
$(\theta(4),\lambda(41))$	0.857 (0.000)	0.862 (0.000)	0.923 (0.000)	0.950 (0.000)	0.964 (0.000)
	13.554 (0.013)	13.694 (0.009)	14.724 (0.010)	15.792 (0.008)	16.856 (0.007)
	0.860 (0.004)	0.862 (0.004)	0.918 (0.003)	0.954 (0.002)	0.967 (0.002)
$(\theta(4),\lambda(42))$	0.883 (0.000)	0.883 (0.000)	0.935 (0.000)	0.957 (0.000)	0.969 (0.000)
	13.553 (0.012)	13.561 (0.012)	14.706 (0.010)	15.817 (0.008)	16.867 (0.007)

We note the following from the numerical results. The larger λ-values obtained from approach 1 usually result in slightly longer tests than those resulting from the smaller values associated with approach 2. The final predictive estimate of field success tends to be slightly conservative (i.e. an underestimate) on the average for approach 1, but tends to be slightly optimistic (i.e. an overestimate) for approach 2. The latter is not surprising since approach 2 adopts priors which imply an overestimate of the initial probability of field success. That said, the results give encouraging evidence of operating characteristics which are robust to the choice of lambda, especially so when c is small. One particular point to note is that for case $\theta(2)$, the characteristics of the Bayes rules when $c = 0.04$ are

very comparable to those obtained from the "3 successes in a row" stopping rule. The reader is referred to Figures 4 and 5 above.

5.2 Stagewise overvariability

Appendix A and Section 4 argue for a modification of (5.2) by the representation of extra-Binomial stage-to-stage variability in the conditional model. The proposal is to replace (5.2) by

$$\tilde{q}_i(d) = E\left(\theta^d\right) \quad \text{with } \theta \sim G_i, \quad 1 \le i \le S . \tag{5.13}$$

Recall that the Binomial model (5.2) was required for the simple structures above based upon sufficient statistics (**S**, **F**). We conclude that, with the more general (5.13), the posterior distribution $\Pi^i(t)$ will depend upon the entire history $H_{it} = \{x_{i1}, x_{i2}, \ldots, x_{it}\}$, 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 lost 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 $D(0)$ defects initially with associated (prior) distribution $\Pi = \{\Pi_d, d \ge 0\}$. The conditional model is $\tilde{q}(d)$, with $\tilde{p}(d) = 1 - \tilde{q}(d)$, $d \ge 0$. 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, $2 \le j \le \varphi$, is the number of successes between failures $(j-1)$ and j. We write $\{\sigma_1, \sigma_2, \ldots, \sigma_{\varphi+1}, \varphi\}$ for this data configuration.

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

$$\Pi_d(t) \equiv \Pi_{d|\sigma_1,\sigma_2,\ldots,\sigma_{\varphi+1},\varphi}$$
$$\propto \Pi_{d+\varphi}\left\{\prod_{j=1}^{\varphi}\tilde{p}(d+j)\right\}\left[\prod_{k=1}^{\varphi+1}\left\{\tilde{q}(d+\varphi+1-k)\right\}^{\sigma_k}\right]. \tag{5.14}$$

Further simplification results in the special case

$$\tilde{q}(d) = \frac{1}{d+1}$$

which results from taking $\theta \sim U[0,1]$ in (5.13). See (A.3) in Appendix A. The posterior distribution in (5.14) then becomes

$$\Pi_{d|\sigma_1,\sigma_2,\ldots,\sigma_{\varphi+1},\varphi} \propto \Pi_{d+\varphi}\left(\frac{d+1}{d+\varphi+1}\right)\left\{\prod_{k=1}^{\varphi+1}\left(\frac{1}{d+\varphi+2-k}\right)^{\sigma_k}\right\}. \tag{5.15}$$

We now perform some calculations which shed light upon the nature of updating and reliability growth in this 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 $\{\sigma_1, \sigma_2, \ldots, \sigma_{\varphi+1}, \varphi\} \equiv \{\sigma, \varphi\}$ and $\{\sigma'_1,\sigma'_2,\ldots,\sigma'_{\varphi+1},\varphi\} \equiv \{\sigma',\varphi\}$.

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 (5.2), the posterior distributions for the two sequences would be identical. This is no longer the case.

We now generalize the material around (5.8) by writing $Q_i(\sigma^i, \varphi^i)\}$ for the predictive probability of field success for subsystem i following data (σ^i, φ^i), $1 \le i \le S$, where

$$Q_i\left\{\left(\sigma^i,\varphi^i\right)\right\} = \sum_{d\ge 0} \Pi^i_{d\left|\sigma^i,\varphi^i\right.} \tilde{Q}_i(d), \quad 1 \le i \le S. \tag{5.16}$$

The corresponding predictive probability for the system as a whole is

$$Q\left\{\left(\sigma^1,\varphi^1\right),\left(\sigma^2,\varphi^2\right),\ldots,\left(\sigma^S,\varphi^S\right)\right\} = \prod_{i=1}^{S} Q_i\left\{\left(\sigma^i,\varphi^i\right)\right\}. \tag{5.17}$$

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

Theorem. For any choices of prior distribution Π^i and conditional model (5.13) for which $\tilde{q}_i(1) = E_{G_i}(\theta) < 1$, the following hold:

(1) If $\{\sigma^i, \varphi^i\}$ dominates $\{\sigma'^i, \varphi^i\}$ then $\Pi^i_{\sigma^i,\varphi^i}$ is stochastically smaller than $\Pi^i_{\sigma'^i,\varphi^i}$;

(2) If the sequence $\{Q^i(d), d \ge 0\}$ is non-decreasing and $\{\sigma^i, \varphi^i\}$ dominates $\{\sigma'^i, \varphi^i\}$, $1 \le i \le S$, then

$$Q_i\left\{\left(\sigma^i,\varphi^i\right)\right\} \ge Q_i\left\{\left(\sigma'^i,\varphi^i\right)\right\}, \quad 1 \le i \le S,$$

and hence

$$Q\left\{\left(\sigma^1,\varphi^1\right),\left(\sigma^2,\varphi^2\right),\ldots,\left(\sigma^S,\varphi^S\right)\right\} \ge Q\left\{\left(\sigma'^1,\varphi^1\right),\left(\sigma'^2,\varphi^2\right),\ldots,\left(\sigma'^S,\varphi^S\right)\right\};$$

(3) If the sequence $\{Q^i(d), d \ge 0\}$ is non-decreasing then $Q_i\{(\sigma^i, \varphi^i)\}$ is non-decreasing in each σ^i_j, $1 \le j \le \varphi^i + 1$, $1 \le i \le S$, as is $Q\{(\sigma^1, \varphi^1), (\sigma^2, \varphi^2), \ldots, (\sigma^S, \varphi^S)\}$.

(4) If $\Pi^i_0 > 0$, $1 \le i \le S$, then during a run of r successes the predictive probability of field success approaches 1 at a geometric rate in r.

Proof.

(1) Let j be such that σ'^i_j and $1 \le j \le \varphi^i$. Consider configuration $\left\{\sigma'^i + \mathbf{1}^{j+1} - \mathbf{1}^j\right\}$. Direct application of (5.14) shows that

$$\begin{aligned} \frac{\Pi^i_{d|\sigma'^i,\varphi^i}}{\Pi^i_{d|\sigma'^i+\mathbf{1}^{j+1}-\mathbf{1}^j,\varphi^i}} &= K_i\left(\sigma'^i,\varphi^i,j\right)\frac{\tilde{q}_i\left(d+\varphi^i+1-j\right)}{\tilde{q}_i\left(d+\varphi^i-j\right)} \\ &= K_i\left(\sigma'^i,\varphi^i,j\right)\frac{E_{G_i}\left(\theta^{d+\varphi^i+1-j}\right)}{E_{G_i}\left(\theta^{d+\varphi^i-j}\right)}, \quad d \ge 0, \end{aligned} \tag{5.18}$$

for some constant $K_i\left(\sigma'^i,\varphi^i,j\right)$. But since distribution G_i has support contained in [0,1] it is straightforward to show that $\left\{E_{G_i}\left(\theta^{d+1}\right)/E_{G_i}\left(\theta^d\right), d \ge 0\right\}$ is a non-decreasing sequence. It follows immediately from (5.18) that the distribution $\Pi^i_{\sigma'^i+\mathbf{1}^{j+1}-\mathbf{1}^j,\varphi^i}$ is smaller than $\Pi^i_{\sigma'^i,\varphi^i}$ 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 $\left(\tilde{\sigma}^i,\varphi^i\right) \rightarrow \left(\tilde{\sigma}^i + \mathbf{1}^{j+1} - \mathbf{1}^j, \varphi^i\right)$ 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 (5.14) and is omitted.

(4) A run of r successes means that each subsystem data configuration is $(r,0)$. By (5.14) we have that

$$Q_i\left\{(r,0)\right\} \ge \Pi^i_{0|r,0} = \frac{\Pi^i_0}{\sum_{d\ge 0}\Pi^i_d\left\{\tilde{q}_i(d)\right\}^r} \ge \frac{\Pi^i_0}{\Pi^i_0 + \left\{1-\Pi^i_0\right\}\left\{\tilde{q}_i(1)\right\}^r}, \tag{5.19}$$

where inequality (5.19) utilizes the decreasing nature of the sequence $\left\{E_{G_i}\left(\theta^d\right), d \ge 0\right\} \equiv \left\{\tilde{q}_i(d), d \ge 0\right\}$. From (5.19) we deduce that that predictive probability of field success for the whole system is

$$Q\{(r,0)\} = \prod_{i=1}^{S} Q_i\{(r,0)\} \geq \prod_{i=1}^{S}\left[1 - \frac{\left(1-\Pi_0^i\right)}{\Pi_0^i}\{\tilde{q}_i(1)\}^r\right]$$

$$\geq 1 - \sum_{i=1}^{S} \frac{\left(1-\Pi_0^i\right)}{\Pi_0^i}\{\tilde{q}_i(1)\}^r$$

and the result is a straightforward consequence.

The conclusions of the above result are strongly suggestive that runs tests, while not being Bayes optimal in the formal sense above, should nevertheless provide simple and effective designs for a range of reasonable cost criteria. We discuss prior analyses of such tests later.

5.3 Bayes confidence regions

A natural focus for inference following testing is the unknown parameter p(system survival in the field). Suppose that, as in (4.4), this takes the value $\prod_{i=1}^{S} \tilde{Q}_i(d_i)$ when the (unknown) number of defectives remaining in subsystem i following testing is d_i, $1 \leq i \leq S$. The case $\tilde{Q}_i(d_i) = Q^{d_i}$, $1 \leq i \leq S$, and Q is a constant is particularly simple and we consider this first. In this model the probability of field survival is $Q^{\sum_{i=1}^{S} d_i}$.

Let $\tilde{\Pi}^i$ be the posterior distribution for the number of defective modes remaining in subsystem i following testing. The $\tilde{\Pi}^i$, $1 \leq i \leq S$, yield $\tilde{\Pi}$, the posterior distribution for the number of defective modes remaining in the entire system following testing. For given $\alpha > 0$, let $D(\alpha)$ be given by

$$D(\alpha) = \min\left(d; \sum_{n=0}^{d} \tilde{\Pi}_n \geq 1-\alpha\right),$$

then $\{1, Q, \ldots, Q^{D(\alpha)}\}$ is a $100(1-\alpha)\%$ Bayes confidence region for the parameter of interest.

In general, we need to work with an ordering of the quantities $\{\prod_{i=1}^{S}\tilde{Q}_i(d_i),\ d_1 \geq 0,$ $d_2 \geq 0, \ldots, d_S \geq 0\}$. Call the (ordered) members of the collection $1 = \tilde{Q}_0 \geq \tilde{Q}_1 \geq \tilde{Q}_2 \geq \ldots$ with

$$\tilde{\mathbf{d}}_r = \left\{\mathbf{d}; \prod_{i=1}^{S}\tilde{Q}_i(d_i) = \tilde{Q}_r\right\}$$

and

$$\tilde{p}_r = \sum_{\mathbf{d}\in\tilde{\mathbf{d}}_r}\left\{\prod_{i=1}^{S}\tilde{\Pi}^i_{d_i}\right\}$$

for the corresponding posterior probability. For given $\alpha > 0$, let $r(\alpha)$ be given by

$$r(\alpha) = \min\left(r; \sum_{n=0}^{r}\tilde{p}_n \geq 1-\alpha\right),$$

then $\left\{1, \tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_{r(\alpha)}\right\}$ is a 100(1 – α)% Bayes confidence region for p(system survival in the field).

5.4 Prior analysis of test designs

As in Section 4, proposed test designs may be assessed by means of a prior analysis (i.e. in advance of the tests) focusing on such key measures as the mean p(system survival in the field) following the test, the mean time to the conclusion of testing, and the probability that the field probability of success is greater than $1 - \alpha$. From a Bayesian viewpoint, the appropriate measures will be expectations taken with respect to the prior distributions Π^i, $1 \leq i \leq S$. Suppose that the $p_r(d_1, d_2, \ldots, d_S)$ are available for $d_i \geq 0$, $1 \leq i \leq S$, by the computations described in Section 4.2. Then

$$\sum_{\mathbf{d}}\left\{\prod_{i=1}^{S}\Pi^i(d_i)\right\}p_r(d_1, d_2 \ldots . d_S) \tag{5.20}$$

is the appropriate measure of say, the mean p(system survival in the field) following a "run of r" test. The summation in (5.20) is over all $d_i \geq 0$, $1 \leq i \leq S$.

In the case of very diffuse priors, implementation of (5.20) may be computationally expensive. Simpler alternatives exist for some of the specially structured models described at the beginning of this section. Consider, for example, a situation in which $\Pi^i \equiv E(\lambda_i, \theta_i)$, $1 \le i \le S$, and we have the conditional Binomial model of (5.2). From (5.9) we conclude that, since $\phi_i = 0$ for this choice of prior, we have

$$\sum_{d\ge 0}\left(1-\theta_i^d\right)\Pi_d^i = 1-\lambda_i, \quad 1\le i\le S,$$

for the unconditional probability of subsystem i success initially. When we further assume that the p(system survival in the field) takes the conditional form $\prod_{i=1}^{S}\theta_i^{d_i}$, we then have for the mean probability of system field success

$$\sum_{\mathbf{d}}\left\{\prod_{i=1}^{S}\Pi^i(d_i)\right\}p_r(d_1,d_2.....d_S) \equiv \tilde{Q}_r(\lambda_1,\lambda_2,\ldots,\lambda_S)$$

$$= \underbrace{\prod_{t=1}^{r}\left\{\prod_{i=1}^{S}\left(1-\lambda_i\theta_i^{t-1}\right)\right\}\prod_{i=1}^{S}\left(1-\lambda_i\theta_i^{r}\right)}_{\text{Run of } r \text{ successful tests occurs before any test failures}}$$

$$+\underbrace{\sum_{t=1}^{r}\sum_{i=1}^{S}\prod_{t'=1}^{t-1}\left\{\prod_{i=1}^{S}\left(1-\lambda_i\theta_i^{t'-1}\right)\right\}\left\{\prod_{j=1}^{i-1}\left(1-\lambda_j\theta_j^{t-1}\right)\right\}\lambda_i\theta_i^{t-1}\,\tilde{Q}_r\left(\lambda_1\theta_1^{t},\ldots,\lambda_{i-1}\theta_{i-1}^{t},\lambda_i\theta_i^{t-1},\ldots,\lambda_S\theta_S^{t-1}\right)}_{\text{Start over after a failure at stage } i \text{ during test } t\le r}$$

and $\tilde{Q}_r(0,0,\ldots,0) = 1$.

Numerical example.

Table 2 reports results from a numerical study of the probability of system field success after a test, which ends with r successes in a row. The system consists of 4 stages. Given d_s defects in stage s, $s = 1, \ldots, 4$, the conditional probability that the system passes one test is $\prod_{s=1}^{4} q_s(d_s)$ where

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

with θ_s having a beta distribution. Two cases of randomized θ_s are considered. In case A, each theta is drawn from a uniform distribution 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

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

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 distributions with the means noted in the table. There are 25 replications for each case. Displayed are the mean of the mean probability of system field success, the means of the probabilities that the probability system field success after the test is greater than or equal to 0.7, 0.8, 0.9, and 0.95, and the mean of the mean number of tests required to obtain a run of r successes. The standard errors of the means appear in parentheses underneath the means.

The distribution of the θ_s, s = 1, ..., 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 3 successes is smaller than for the case of uniformly distributed θ_s, s = 1, ..., 4.

The initial mean number of defects in each stage also affects the probability of field success. The case with 5 defects in stage 4 has the smallest mean of the mean probability of field success after a test. The mean of the mean probabilities of field success after a test until there is a run of 7 successes in a row is 0.66 for this case.

The mean of the mean 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 probabilities of field success are higher. It is important to design tests so that they represent field conditions as closely as possible.

Table 2

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 (0.04)	0.28 (0.04)	0.13 (0.02)	0.13 (0.02)	5	0.8	B	17.32 (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	20.22 (0.95)

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 $s + 1, \ldots, S$ 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 propose analytical procedures to calculate the mean probability of field system survival after successful completion of a runs test, the distribution of the probability of system field survival after a successful runs test, and the mean 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 faults 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. This suggests that it is important to design operational tests so that the test mimics field operation of the system as closely as possible.

The procedures of this paper have been programmed in Visual Basic and Excel. The software is available from PAJ. Exercise of such software can provide guidance to test planners and analysts.

Acknowledgement

The authors acknowledge with thanks the simulation implementation of the Bayes test procedures by Michael Dugas; the results of the simulation are displayed in Table 1.

Bibliography

Ascher, H., and Feingold, H., *Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes, Lecture Notes in Statistics*, **7**, Marcel Dekker, New York, 1984.

Balaban, H.S., "Reliability Growth Models," *Journal of Environmental Sciences*, **21**, 11-18 (Jan/Feb 1978).

Barlow, R.E., and Scheuer, E.M., "Reliability Growth During a Development Testing Program," *Technometrics*, **8**, 53-60 (Feb 1966).

Barr, D.R., "A Class of General Reliability Growth Models," *Operations Research*, **18**, 52-65 (1970).

Benkherouf, L. and Bather, J.A., "Oil exploration: sequential decisions in the face of uncertainty," *Journal of Applied Probability*, **25**, 529-543 (1988).

Benkherouf, L., Glazebrook, K.D., and Owen, R.W., "Gittins indices and oil exploration," *Journal of the Royal Statistical Society*, **B54**, 229-241 (1992).

Bhattacharyya, G.K., Fries, A., and Johnson, R.A., "Properties of Continuous Analog Estimators for a Discrete Reliability-Growth Model," *IEEE Transactions on Reliability*, **38**, 373-378 (Aug 1989).

Block, Henry W. and Savits, Thomas H., "Burn-In," *Statistical Science*, **12/1**, 1-13 (1997).

Bogdonavicius, V. and Nikulin, M.S., "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).

Calabria, R., Guida, M., and Pulcini, G., "A Bayes Procedure for Estimation of Current System Reliability," *IEEE Transactions on Reliability*, **41**, 616-621 (Dec 1992).

Chernoff, H. and Ray, S.N., "A Bayes Sequential Sampling Inspection Plan," *Annals of Mathematical Statistics*, **36/5**, (October 1965).

Chernoff, H., "Sequential Models for Clinical Trials," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, 1966.

Easterling, Robert G., Mazumdar, Mainak, Spencer, Floyd W., and Diegert, Kathleen V., "System-based component-test plans and operating characteristics: binomial data," *TECHNOMETRICS*, **33/3**, 287-298 (August 1991).

Feller, W. *An Introduction to Probability Theory and Its Applications*, Volume II. John Wiley & Sons, Inc., New York, 1966.

Fries, A. and Maillart, L.M. "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).

Fries, A., "Discrete Reliability-Growth Models Based on a Learning-Curve Property," *IEEE Transactions on Reliability*, **42**, 303-306 (June 1993).

Gaver, D.P. and Jacobs, P.A., "Methodology for an operationally-based test length decision," *IIE Transactions*, **30**, 1129-1134 (1998).

Gaver, D.P. and Jacobs, P.A., "Testing or Fault-Finding for Reliability Growth: A Missile Destructive-Test Example," *Naval Research Logistics*, **44**, 623-637 (1997).

Gaver, D.P. and Jacobs, P.A., "WEBSIGHT: An internet tool for uncertainty analysis in operational testing," American Statistical Association Joint Annual Meetings, Baltimore, MD, August 8-12, 1999.

Gittins, J.C., *Multi-armed Bandit Allocation Indices*, New York: Wiley, 1989.

Glazebrook, K.D., "On a family of prior distributions for a class of Bayesian search models," *Advances in Applied Probability*, **25**, 714-716 (1993).

Gross, A.J., and Kamins, M., "Reliability Assessment in the Presence of Reliability Growth," *Annals of Assurance Sciences: 1968 Symposium on Reliability*, 406-416 (1968).

Jayachadran, T., and Moore, L.R., "A Comparison of Reliability Growth Models," *IEEE Transactions on Reliability*, **R-25**, 49-51 (Apr 1976).

Johnson, N.L., Katz, S., and Balakrishnan, N., *Continuous Univariate Distributions*, Vol. 1, Second Edition, New York: Wiley, 1994.

Lynn, N.J. and Singpurwalla, N.D., "Comment: 'Burn-In' makes us feel good," *Statistical Science*, **12**, 13-19 (1997).

Manor, G. and Kress, M., "Optimality of the greedy shooting strategy in the presence of incomplete damage information," *Naval Research Logistics*, **44**, 613-622 (1997).

Mazzuchi, T.A., and Soyer, R., "A Bayes Method for Assessing Product-Reliability During Development Testing," *IEEE Transactions on Reliability*, **42**, 503-510 (Sep 1993).

Olsen, D.E., "Estimating Reliability Growth," *IEEE Transactions on Reliability*, **R-26**, 50-53 (Apr 1977).

Pollock, S.M., "A Bayesian Reliability Growth Model," *IEEE Transactions on Reliability*, **R-17**, 187-198 (Dec 1968).

Seglie, E.A. (1998) "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.

Sen, A., Basu, S., and Banerjee, M., "Statistical Analysis of Life-Data with Masked Cause-of-Failure," *Handbook of Statistics: Reliability* (C.R. Rao, N. Balakrishnan, eds.) (2000, to appear).

Woods, W.M., "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., 200-204 (1990).

Yang, M.C., Wackerly, D.D., and Rosalsky, A., "Optimal stopping rules in proof-reading," *J. Appl. Prob.* **19**, 723-729 (1982).

APPENDIX A
Models for Stage-wise Over/Extra Variability

Generalize the initial binomial stagewise (sub)model by randomizing θ_s: replace θ_s by the random variable θ_s and replace $\theta_s^{d_s}$ by $E\left[\theta_s^{d_s}\right] = q_s(d_s)$.

Beta Mixing:

$$E\left[\theta_s^{d_s}\right] \equiv q_s(d_s) = \frac{\Gamma(a_s+b_s)}{\Gamma(a_s)\Gamma(b_s)}\int_0^1 x^{a_s-1}(1-x)^{b_s-1}(1-x)^{d_s}\,dx = \frac{\Gamma(a_s+b_s)}{\Gamma(b_s)}\frac{\Gamma(b_s+d_s)}{\Gamma(a_s+b_s+d_s)} \quad \text{(A.1)}$$

One reasonable normalization is

$$\theta_s = \frac{b_s}{a_s+b_s} \quad \text{(A.2)}$$

where θ_s is the original "deterministic" survival probability, i.e. put $q_s(1) = \theta_s$.

The above model simultaneously chooses the *same* random value for each defect in a stage for each visit to the stage, and is independent between stages. For a uniform distribution ($a_s = b_s = 1$),

$$q_s(d_s) = \frac{1}{1+d_s} \quad \text{(A.3)}$$

while the corresponding non-mixed version is $\left(\frac{1}{2}\right)^{d_s}$, the latter decreasing much more rapidly than the former.

Alternative ("Transform") Mixing:

If

$$q_s(d_s) = E\left[\theta_s^{d_s}\right] = E\left[e^{-d_s(-\ln\theta_s)}\right], \quad \text{(A.4)}$$

the Laplace transform of the positive random variable $(-\ln\theta_s)$; here are some tractable possibilities:

Gamma Mixing:

The Laplace transform of $Y \sim \text{Gam}(\beta = \text{shape}, \mu = \text{mean})$ is

$$E\left[e^{-\xi Y}\right] = \left(1 + \frac{\mu\xi}{\beta}\right)^{-\beta} \tag{A.5}$$

so if $Y = -\ln\theta$, put $\xi = d$ to find

$$q(d) = \frac{1}{\left(1 + \frac{\mu d}{\beta}\right)^{\beta}} \tag{A.6}$$

If $\mu = \beta = 1$ the result equals the uniform (Beta) result, but the Gamma result above is more flexible. Normalizing at $d = 1$,

$$q(1) = \frac{1}{\left(1 + \frac{\mu}{\beta}\right)^{\beta}} = \theta. \tag{A.7}$$

If β is an optional tuning parameter,

$$\mu = \beta\left(\theta^{-1/\beta} - 1\right). \tag{A.8}$$

Stable Law Mixing:

The Laplace transform of a positive stable law (see Feller, 1966) is

$$E\left[e^{-\xi Y}\right] = e^{-(\alpha\xi)^{\beta}}, \quad \text{for } \alpha > 0 \text{ and } 0 < \beta < 1 \tag{A.9}$$

If $Y = -\ln\theta$, then $q(d) = e^{-(\alpha d)^{\beta}}$. Normalize at $d = 1$ to get

$$q(d) = \theta^{d^{\beta}} \tag{A.10}$$

where θ is the "deterministic" probability of defect survival.

Inverse Gaussian (IG) Mixing (see Johnson, Kotz, and Balakrishnan, 1994):

The IG distribution is that of the first-passage time of a Brownian motion with drift. If $Y = -\ln\theta$ then

$$q(d) = E\left[\theta^d\right] = E\left[e^{-dY}\right] = \exp\left[-\frac{1}{c}\left\{(1+2c\mu d)^{\frac{1}{2}} - 1\right\}\right] \qquad \text{(A.11)}$$

where

$$\mu = E[-\ln\theta], \quad c = \frac{\mathrm{Var}[-\ln\theta]}{\mu} \qquad \text{(A.12)}$$

the latter being the coefficient of variation of ($-\ln\theta$). $q(d) = e^{-\mu d}$ if $c \to 0$. The $q(d)$ depends on the tuning parameter c.

APPENDIX B
Effects of Test-to-Test Variability

Let $\boldsymbol{T}(d_1, d_2, \ldots, d_j, \ldots, d_S; \underline{\varepsilon})$ denote the random number of tests to achieve a run of r successes for the first time, conditional d_i defects being initially present in stage i, $i = 1, \ldots, S$ and on the environmental test-specific random variables $\underline{\varepsilon}$; these latter are assumed positive, independently sampled, and held fixed for each entire test; they thus represent test-to-test variability, which can be random (as here), but also deterministic, explanatory/regression variables. It can be seen that, conditional on $\underline{\varepsilon}$ components,

$$\boldsymbol{T}(d_1, d_2, \ldots, d_j, \ldots, d_S; \underline{\varepsilon})$$

$$= n + \boldsymbol{T}(d_1, d_2, \ldots, d_j - 1, \ldots, d_S; \underline{\varepsilon}') \text{ for } n = 1, 2, \ldots, r; j = 1, 2, \ldots, S, \text{ with probability}$$

$$\prod_{i=1}^{S} \left(\tilde{q}_i(d_i)\right)^{\varepsilon_1} \prod_{i=1}^{S} \left(\tilde{q}_i(d_i)\right)^{\varepsilon_2} \cdots \prod_{i=1}^{S} \left(\tilde{q}_i(d_i)\right)^{\varepsilon_{n-1}} \times \prod_{i=1}^{j-1} \left(\tilde{q}_i(d_i)\right)^{\varepsilon_n} \left(1 - \left(\tilde{q}_j(d_j)\right)^{\varepsilon_n}\right);$$

$$= r \text{ with probability } \prod_{i=1}^{S} \left(\tilde{q}_i(d_i)\right)^{\varepsilon_1} \prod_{i=1}^{S} \left(\tilde{q}_i(d_i)\right)^{\varepsilon_2} \cdots \prod_{i=1}^{S} \left(\tilde{q}_i(d_i)\right)^{\varepsilon_r}. \qquad \text{(B.1)}$$

The next steps lead to finding the mean time to first attain a run of r test successes, thus stopping the overall test.

First, take the expectation, conditional on $\underline{\varepsilon}$:

$$E\left[\boldsymbol{T}(d_1, d_2, \ldots, d_S; \underline{\varepsilon})\right]$$

$$= \underbrace{\sum_{n=1}^{r} \sum_{j=1}^{S} \left\{ n + E\left[\boldsymbol{T}(d_1, d_2, \ldots, d_j - 1, \ldots, d_S; \underline{\varepsilon}')\right] \right\} \prod_{k=1}^{n-1} \prod_{i=1}^{S} \left(q_i(d_i)\right)^{\varepsilon_k} \cdot \prod_{i=1}^{j-1} \left(q_i(d_i)\right)^{\varepsilon_n} \left(1 - \left(q_j(d_j)\right)^{\varepsilon_n}\right)}_{\substack{\text{The conditional (on } \underline{\varepsilon}\text{) expected time to run of } r\text{, when there is no run of } r \text{ in first } r \\ \text{(run-breaker occurs at Stage } j\text{; next test starts over with } d_j - 1 \text{ defects in Stage } j\text{)}}}$$

$$+ \underbrace{r \prod_{i=1}^{S} \left(q_i(d_i)\right)^{\varepsilon_1} \prod_{i=1}^{S} \left(q_i(d_i)\right)^{\varepsilon_2} \ldots \prod_{i=1}^{S} \left(q_i(d_i)\right)^{\varepsilon_r}}_{\substack{\text{The conditional expected value of run length,} \\ \text{when the run occurs on the first } r \text{ tests}}} \qquad \text{(B.2)}$$

Next, remove the condition on $\underline{\varepsilon}$, noting that the $\underline{\varepsilon}'$ appearing in $\boldsymbol{T}(d_1, d_2, \ldots, d_j - 1, \ldots, d_S; \underline{\varepsilon}')$ refers to future tests, and is here assumed independent on all before; this is a

plausible convenience but not a necessity. Taking expectations or removing the $\underline{\varepsilon}$-condition under the iid assumption yields

$$r_r(d_1,d_2,\ldots,d_S) \equiv E_\varepsilon E\left[T(d_1,d_2,\ldots,d_S;\underline{\varepsilon})\right]$$

$$= \sum_{n=1}^{r}\sum_{j=1}^{S}\left\{n + r(d_1,\ldots,d_j-1,\ldots,d_S)\right\}\left(M_S(d_1,d_2,\ldots,d_S)\right)^{n-1}\left(M_{j-1}(d_1,d_2,\ldots,d_S) - M_j(d_1,d_2,\ldots,d_S)\right)$$

$$+ r\left(M_S(d_1,d_2,\ldots,d_S)\right)^r \qquad \text{(B.3)}$$

where

$$M_{j-1}(d_1,d_2,\ldots,d_S) = E_\varepsilon\left[\prod_{i=1}^{j-1}\left(q_i(d_i)\right)^\varepsilon\right]$$

$$= E_\varepsilon\left[e^{-\varepsilon\left(-\sum_{i=1}^{j-1}\ln q_i(d_i)\right)}\right] \qquad \text{(B.4)}$$

and

$$M_S(d_1,d_2,\ldots,d_S) = E_\varepsilon\left[\prod_{i=1}^{S}\left(q_i(d_i)\right)\right]$$

$$= E_\varepsilon\left[e^{-\varepsilon\left(-\sum_{i=1}^{S}\ln q_i(d_i)\right)}\right] \qquad \text{(B.5)}$$

The latter expressions can be evaluated in terms of the Laplace transform of the ε-distribution. Many such transforms of distributions are simple and explicit; see Appendix A for examples.

Let $\boldsymbol{R}(d_1, d_2, \ldots, d_S; \underline{\varepsilon})$ denote the random probability of success after passing a test stopped after a run of r successes, again conditional on test-specific environmental random variables $\underline{\varepsilon}$; $\varepsilon(f)$ refers to *field* environments, which may differ from those of the tests.

$$\boldsymbol{R}(d_1,d_2,\ldots,d_S;\underline{\varepsilon}) = \boldsymbol{R}(d_1,d_2,\ldots,d_i-1,\ldots,d_S;\underline{\varepsilon}')$$

for $n = 1, 2, \ldots, r-1$ and $i = 1, 2, \ldots, S$ with probability

$$\underbrace{\prod_{i=1}^{S}\left(\tilde{q}_i(d_i)\right)^{\varepsilon_1}\prod_{i=1}^{S}\left(\tilde{q}_i(d_i)\right)^{\varepsilon_2}\cdots\prod_{i=1}^{S}\left(\tilde{q}_i(d_i)\right)^{\varepsilon_{n-1}}\times\prod_{j=1}^{i-1}\left(\tilde{q}_j(d_j)\right)^{\varepsilon_n}\left(1-\left(\tilde{q}_i(d_i)\right)^{\varepsilon_n}\right)}_{\text{failure before run of } r\text{; start over}}$$

$$=\underbrace{\prod_{i=1}^{S}\left(\tilde{Q}_i(d_i)\right)^{\varepsilon_i(f)}}_{\text{no failure during run of } r} \quad \text{with probability } \prod_{i=1}^{S}\left(\tilde{q}_i(d_i)\right)^{\varepsilon_1}\cdots\prod_{i=1}^{S}\left(\tilde{q}_i(d_i)\right)^{\varepsilon_r}$$

Remove conditions on ε_x (iid) and sum, using

$$M_j(d_1,d_2,\ldots,d_S) = E_\varepsilon\left[\exp\left\{-\varepsilon\sum_{i=1}^{j}\left(-\ln\tilde{q}_i(d_i)\right)\right\}\right] \quad j = 1,2,\ldots,S$$

$$M_{S,f}(d_1,d_2,\ldots,d_S) = E_{\varepsilon(f)}\left[\exp\left\{-\varepsilon\sum_{i=1}^{S}\left(-\ln\tilde{Q}_i(d_i)\right)\right\}\right]$$

$$\Pr(d_1,d_2,\ldots,d_S) = E_\varepsilon E\left[\boldsymbol{R}(d_1,d_2,\ldots,d_S;\underline{\varepsilon})\right]$$

$$=\underbrace{\left(M_S(d_1,d_2,\ldots,d_S)\right)^r M_{S,f}(d_1,d_2,\ldots,d_S)}_{\text{run of } r \text{ successful tests before any failures}}+\underbrace{\sum_{n=1}^{r}\left(M_S(d_1,d_2,\ldots,d_S)\right)^{n-1}}_{=\ \underbrace{\frac{1-\left(M_S(d_1,\ldots,d_S)\right)^r}{1-M_S(d_1,\ldots,d_S)}}_{\text{no } r\text{-run during first } r \text{ tests}}}$$

$$\times\underbrace{\sum_{i=1}^{S}\Pr(d_1,d_2,\ldots,d_i-1,\ldots,d_S)\left[M_{i-1}(d_1,d_2,\ldots,d_S)-M_i(d_1,d_2,\ldots,d_S)\right]}_{\textit{start over} \text{ after failure at some stage (i) before achieve run of } r\text{; remove defect, continue}}$$

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 B.1 displays probabilities of system field success for a system that has been tested until there is a run of 5 successes. The testing environmental random variables have a gamma distribution with mean 1 and shape parameter 0.5. The field environmental

random variable has a gamma distribution with mean 1 and shape parameter beta, which has been made widely variable. Note that the smaller beta 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 the quite disparate field conditions, quite insensitive to the distribution of random field effects.

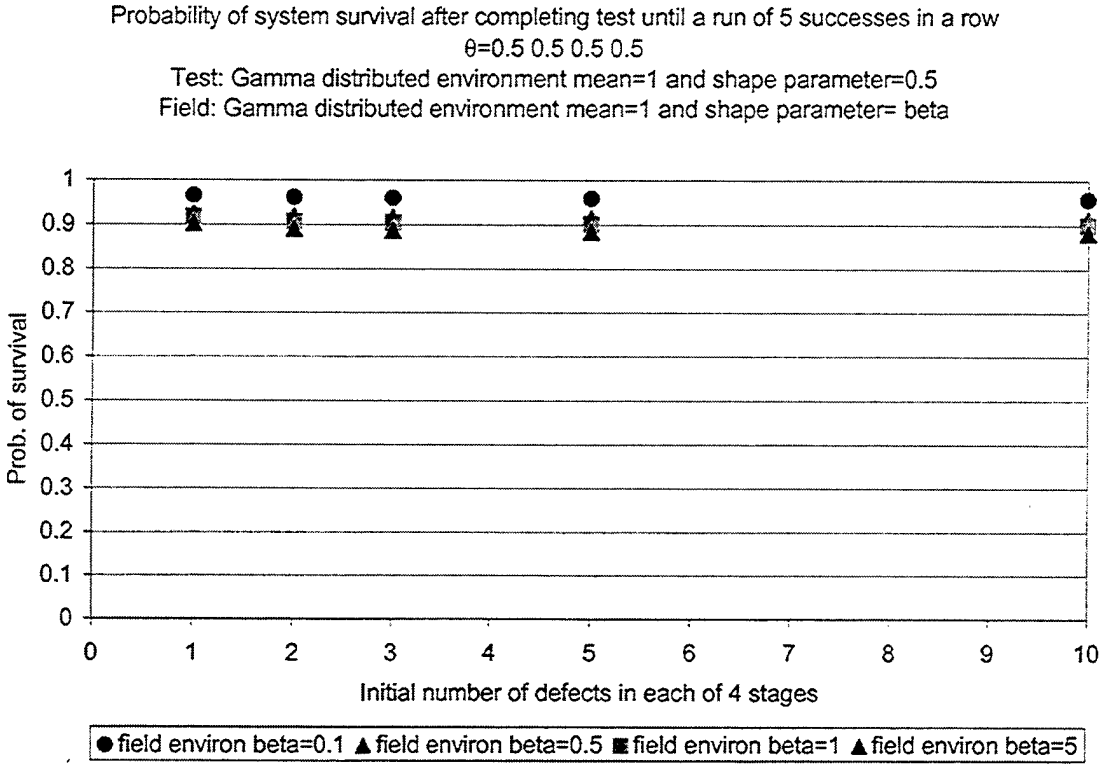

Figure B.1

In contrast to Figure B.1, Figure B.2 (respectively B.3) displays probabilities of field success (respectively, the number of tests to obtain 5 successes in a row) for a system that has again been 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 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. The reason: 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 faults. 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.

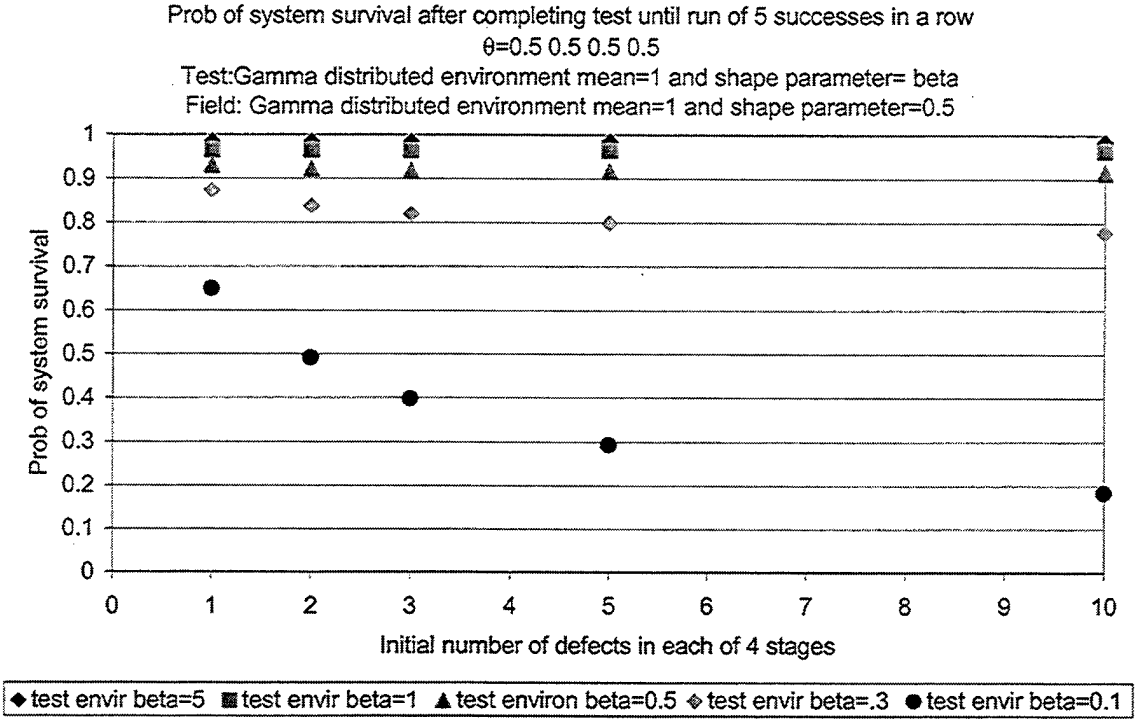

Figure B.2

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. This is only common sense, but quantified.

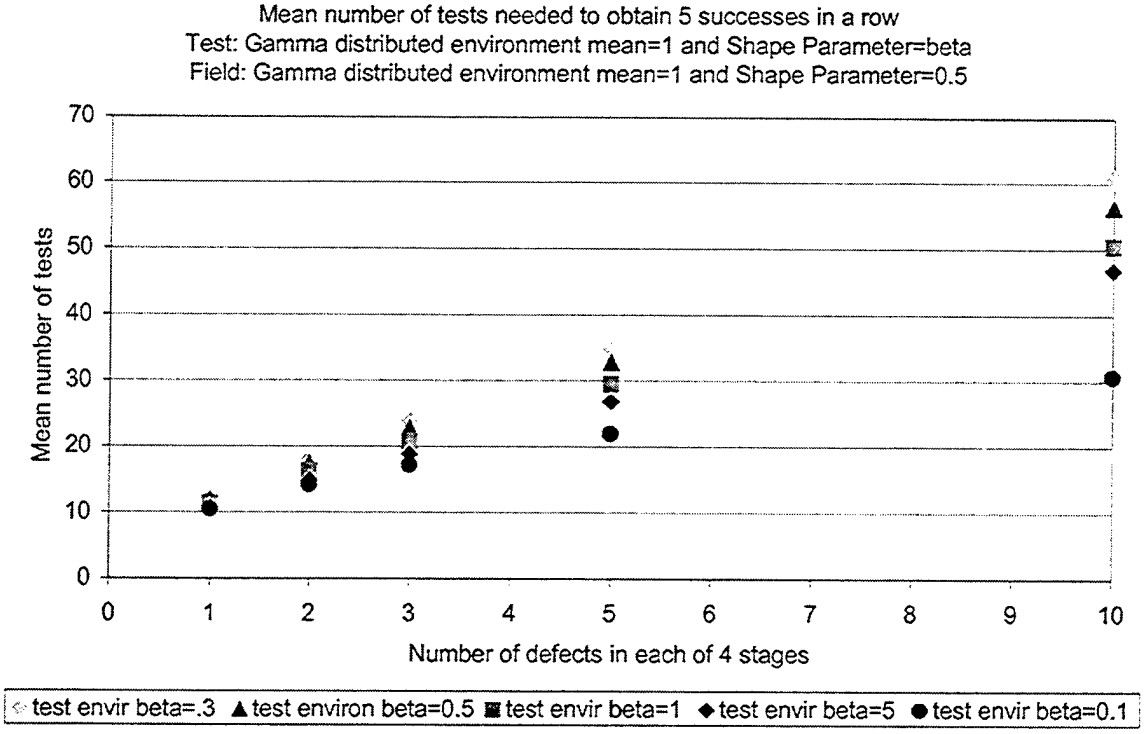
Mean number of tests needed to obtain 5 successes in a row
Test: Gamma distributed environment mean=1 and Shape Parameter=beta
Field: Gamma distributed environment mean=1 and Shape Parameter=0.5
Mean number of tests
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
Number of defects in each of 4 stages
test envir beta=.3
test environ beta=0.5
test envir beta=1
test envir beta=5
test envir beta=0.1

Figure B.3

DISTRIBUTION LIST

1. Research Office (Code 09) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Dudley Knox Library (Code 013) 2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Defense Technical Information Center 2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

4. Therese Bilodeau (Editorial Assistant) 1
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Prof. Donald P. Gaver (Code OR/Gv) 3
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Prof. Patricia A. Jacobs (Code OR/Jc) 25
   Dept of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

7. Prof. Kevin D. Glazebrook 2
   Dept. of Statistics
   Newcastle University
   Newcastle upon Tyne NER17RU
   ENGLAND

8. Dr. Ernest Seglie 1
   Science Director, DOT&E
   3E318 Pentagon
   Washington, DC 20301-1700

9. Mr. Michael Bauman 1
   Director, USA Training & Doctrine Command Analysis Center (TRAC)
   Fort Leavenworth, KS 66027

10. Mr. Kevin Becker 1
    Compaq Computer Corp.
    MS CAC 06-64
    1055 Ridgeview Ct.
    Cupertino, CA 95014

11. Prof. Henry Block 1
    Dept. of Math and Statistics
    University of Pittsburgh
    Pittsburgh, PA 15260

12. Prof. Dan Boger 1
Chair, C3I Academic Group
Naval Postgraduate School
Monterey, CA 93943-5000

13. Dr. Robert L. Carpenter 1
Naval Health Research Center Detachment – Toxicology
2612 Fifth St., Bldg. 433, Area B
Wright-Patterson AFB, OH 45433-7903

14. Center for Naval Analyses 1
4401 Ford Avenue
Alexandria, VA 22302-0268

15. Dr. W. Peter Cherry 1
Vector Research Inc.
P.O. Box 1506
Ann Arbor, MI 48106

16. Mr. Thomas Christie 1
Institute for Defense Analysis
1800 North Beauregard
Alexandria, VA 22311

17. Dr. Michael L. Cohen 1
Study Director
Committee on National Statistics
2101 Constitution Ave.
Washington, DC 20418

18. Prof. Sir David Cox 1
Nuffield College
Oxford OXI INF
ENGLAND

19. Dr. Larry Crow 1
AT&T Bell Laboratories
67 Whippany Road
Whippany, NJ 07981

20. Dr. Siddhartha Dalal 1
Telcordia Technologies
445 South Street
Morristown, NJ 07960-6438

21. Dr. D. F. Daley 1
Statistics Dept. (I.A.S.)
Australian National University
Canberra, A.C.T 2606
AUSTRALIA

22. Dr. Robert Easterling 1
Sandia National Laboratories
MS 0417
Albuquerque, NM 87195

23. USAMSAA 1
ATTN: (ATSD Dr. Paul M. Ellner)
392 Hopkins Road
Aberdeen Proving Ground, MD 21005-5071

24. Dr. Arthur Fries 1
Institute for Defense Analysis
1800 North Beauregard
Alexandria, VA 22311

25. Dr. Andrew Gelman 1
Statistics Dept.
Columbia University
New York, NY 10027

26. Prof. Bernard Harris 1
Dept. of Statistics
University of Wisconsin
610 Walnut Street
Madison, WI 53706

27. Mr. Walter W. Hollis 1
Deputy Under Secretary of the Army (OR)
ATTN: SAUS (OR)
The Pentagon, Room 2E660
Washington, DC 20310-0102

28. Prof. J. B. Kadane 1
Dept. of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

29. Dr. Sallie Keller-McNulty 1
Los Alamos National Laboratory
Group TSA-1
MSF 600
Los Alamos, NM 87545

30. Dr. Jon Kettenring 1
Bellcore
445 South Street
Morris Township, NJ 07962-1910

31. Dr. Moshe Kress 1
CEMA
P.O.B. 2250 (TI)
Haifa 31021
ISRAEL

32. Prof. Guy Latouche 1
University Libre Bruxelles
C.P. 212, Blvd. De Triomphe
Bruxelles B-1050
BELGIUM

33. Dr. A. J. Lawrance 1
Dept. of Mathematics
University of Birmingham
P.O. Box 363
Birmingham B15 2TT
ENGLAND

34. Prof. J. Lehoczky 1
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

35. Dr. Colin Mallows 1
AT&T Bell Telephone Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

36. Dr. Harry Martz 1
Los Alamos National Laboratory
Group TSA-1
MSF 600
Los Alamos, NM 87545

37. Prof. M. Mazumdar 1
Dept. of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA 15235

38. Prof. Ali Mosleh 1
University of Maryland
Dept of Materials & Nuclear Engineering
Rm. 2100 A
Marie Mount Hall
College Park, MD 20742-7531

39. Prof. M. Nikuline 1
UFR MIZS Universite Bordeaux 2
Boite Postale 69
Bordeaux 33076
FRANCE

40. Dr. David H. Olwell 1
1877 N. Marble Ridge Pl.
Tucson, AZ 85715

41. Prof. Steven Pollock 1
Dept of Industrial-Operations Engr.
1831 IOE
University of Michigan
Ann Arbor, MI 48109

42. Prof. Jesse Poore 1
Dept of Computer Science
University of Tennessee
Knoxville, TN 37996

43. Dr. John E. Rolph 1
Information and Operations Management
Univ. of Southern California
School of Business Administration
Los Angeles, CA 90089-1421

44. Prof. M.F. Romalhoto 1
Departamento de Matematica
Instituto Superior Tecnica
Avenida Rovisco Pais
Lisboa 1096
PORTUGAL

45. Prof. Frank Samaniego 1
Statistics Department
University of California
Davis, CA 95616

46. Prof. Sam Saunders 1
PO Box 458
Kirkland, WA 98083-0458

47. Prof. Tom Savits 1
Dept. of Math and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

48. Prof. N. D. Singpurwalla 1
George Washington University
Washington, DC 20052

49. Prof. Duane Steffey 1
Dept. of Math and Computer Science
San Diego State University
5500 Campanile Dr.
San Diego, CA 92182-7720

50. Prof. L.C. Thomas 1
School of Management
University of Southampton
Highfield
Southampton SO17 1BJ
ENGLAND

51. Dr. Michael Tortorella 1
Bell Laboratories
M/S 2L-536
101 Crawford's Corner Rd.
Holmdel, NJ 07733-3030

52. Dr. D. Vere-Jones 1
Dept. of Math
Victoria Univ. of Wellington
P.O. Box 196
Wellington
NEW ZEALAND

53. Dr. L. Wein 1
Operations Research Center, Rm E40-164
Massachusetts Institute of Technology
Cambridge, MA 02139

54. Dr. Gideon Weiss 1
Dept. of Statistics
The University of Haifa
Haifa 31905
ISRAEL

55. Mr. Steven Whitehead 1
c/o Commander, Operational Test & Evaluation Force
7970 Diven Street
Norfolk, VA 23505

56. Dr. Daniel Willard 1
Deputy Under Secretary of the Army (OR)
ATTN: SAUS (OR)
The Pentagon, Room 2E660
Washington, DC 20310-0102

57. Dr. Alyson Wilson 1
Los Alamos National Laboratory
Group TSA-1
MSF 600
Los Alamos, NM 87545