(12) LEVEL II

ORC 78-9
JUNE 1978

"RELIABILITY GROWTH" AS AN ARTIFACT OF RENEWAL TESTING

by

WILLIAM S. JEWELL

OPERATIONS
RESEARCH
CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

9 Research rept.,

6 "RELIABILITY GROWTH" AS AN ARTIFACT OF RENEWAL TESTING

by

10 William S. Jewell
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

15 AFOSR-77-3179 16 2304

17 A5

12 23p.

11 JUNE 1978

14 ORC-78-9

This research was supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-77-3179 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

270 750

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ORC 78-9	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) "RELIABILITY GROWTH" AS AN ARTIFACT OF RENEWAL TESTING		5. TYPE OF REPORT & PERIOD COVERED Research Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William S. Jewell		8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3179
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling AFB, D.C. 20332		12. REPORT DATE June 1978 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Reliability Growth
MTBF Estimators
Renewal Testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

It is shown that observed "reliability growth" may be an artifact of limited-horizon renewal testing. The "growth" of several estimators of failure rate and MTBF with test time is examined for a stationary renewal process.

"RELIABILITY GROWTH" AS AN ARTIFACT OF RENEWAL TESTING

by

William S. Jewell

1. INTRODUCTION

"Reliability growth" refers to the empirically-observed fact that the performance of a complex system shows improvement as the newly-designed system moves into the development program and is tested under actual or simulated operational conditions. This improvement results primarily from the identification and correction of initial design and engineering deficiencies [6], but may also be due to improved familiarity in the field by operating personnel, revised operating and maintenance procedures, etc. Analysis and extrapolation of reliability growth is of critical importance to both the manufacturer and the purchasing agency because of the need to balance (1) the high costs of making engineering changes; (2) the costs and delays in extended testing; (3) the operational support costs of maintenance and replacement; and (4) the need for operational readiness and reliability. Currently, the U.S. Air Force is experimenting with a new procurement procedure, the Reliability Improvement Warrant [3], [8], [12], under which financial incentives are offered to contractors "to design and produce equipment which will have a low failure rate as well as low repair costs after failure due to field/operational use" [12].

Beginning with [14], there have been many models proposed for reliability growth [5], [7]. The two major approaches [11] are: (1) to assume that observable improvements occur only in discrete phases, as major redesign efforts follow testing; (2) to assume that a continuous "learning curve" of reliability improvement applies during the development program. The most successful model of the second type is due to J. T. Duane (see [7]); it assumes that the failure process is time-varying Poisson, with cumulative mean failure rate a decreasing algebraic (Weibull) function of system total operating time (including fixes and replacements).

The purpose of this paper is to suggest that some of the "reliability growth" supposedly observed during testing programs may be an artifact of the parameter chosen for measurement, and the statistic used to estimate that parameter, particularly in short-duration field performance demonstration programs after major design developments have been implemented. Our approach, suggested by I. Shimi, is to assume that the underlying failure process is, in fact, *stable*, and to show that reasonable estimators are *biased* functions of the testing interval length. These results suggest that it may be difficult to separate true reliability growth from the measurement process.

2. TESTING MODEL AND PERFORMANCE ESTIMATORS

We assume that the failure mechanism of the system generates a renewal process, in which successive random intervals between failures, $X_1, X_2, \ldots$ are independent and *identically distributed* samples from a distribution function F, with density $f(x) = dF(x)/dx$, complement $\bar{F} = 1 - F$, and moments $m_j = E\{X^j\}$. Consider a single system placed on *renewal testing* for a total operating time of t hours; during this interval a random number $N(t)$ of failures and instantaneous fixes occur, with the last failure occurring at epoch $Y_{N(t)} = X_1 + X_2 + \cdots + X_{N(t)}$, $Y_{N(t)} \leq t$ $(N(t) > 0)$ (Figure 1).

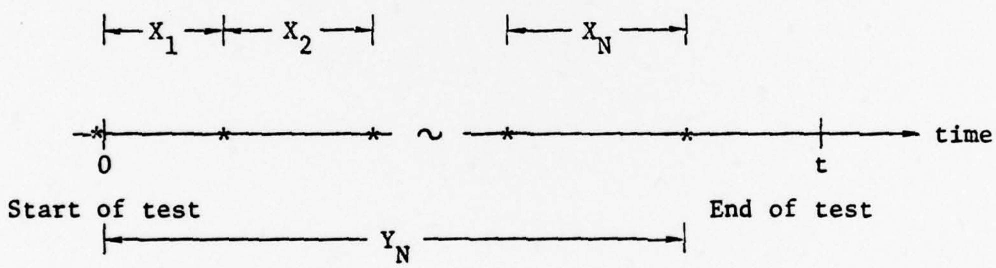

Figure 1

Renewal Testing Scheme

The first performance estimator to be considered might be the *empirical failure rate*

$$(1) \qquad \psi(t) = \frac{N(t)}{t} \qquad (t > 0) .$$

Presumably, if there is "reliability growth," $N(t)$ will be "concave down" [2], or $\psi(t)$ will be decreasing with t . [13].

A more useful performance measure, and the one upon which the Reliability Improvement Warranty is based, is the mean-time-between-failure (MTBF). Depending upon one's statistical sophistication, there are three estimators that suggest themselves from Figure (1):

1. The Naive MTBF Estimator

Since there are $N(t)$ complete samples observed, the sample mean lifetime should be used as MTBF estimator:

$$(2) \qquad \mu_1(t) = \sum_{i=1}^{N(t)} X_i/N(t) = Y_{N(T)}/N(t) \text{ , } (t > 0)$$

assuming that at least one failure occurs in $(0,T]$ $(N(t) > 0)$.

2. The Clever MTBF Estimator

The alert reliability engineer will realize that the residual interval $t - Y_{N(t)}$, represents incomplete sample information that should not be discarded. Further, since exponentially-distributed lifetimes are of important practical interest, and it is known that these lifetimes are "memoryless," one might argue that the incomplete sample is similar to an $(N(t) + 1)^{st}$ sample, so the MTBF estimator should be:

(3) $$\mu_2(t) = \frac{t}{N(t) + 1} \qquad (t > 0)$$

which is defined even if there are no failures.

3. The Sophisticated MTBF Estimator

A sophisticated analyst will have read the literature on renewal testing (see [5] and [9]), and will realize that the maximum likelihood estimator of MTBF in the exponential case is *the total-time-on-test* divided by the number of *complete* samples, or

(4) $$\mu_3(t) = t/N(t) \ , \ (t > 0)$$

again defined only if $N(t) > 0$.

In examining the above estimates, it is clear that $\mu_3(t) = 1/\psi(t)$, $\mu_3(t) > \mu_1(t)$, and $\mu_3(t) > \mu_2(t)$, but that the actual sample values may follow rather complicated paths as functions of the testing interval t .

In what follows we shall take two approaches:

(1) Find the expected value of each estimator. This reflects the mean value of the sample function for a single-system test, and also approximates the sample function if a very large number of systems are on test simultaneously, and the individual estimators are *averaged*;

(2) Secondly, we shall examine whether *pooling* the estimates for many identical systems on test provides any improvement, by looking at the limiting values of these pooled statistics.

Naturally, a more complete analysis would examine the distribution of each statistic with a finite number of systems on test; but, as we shall see, the (undesired) bias of these estimators negates any more detailed analysis.

3. THE FAILURE RATE ESTIMATOR

If the estimator (1) is used, its mean value is:

$$E\{\psi(t)\} = M(t)/t \tag{5}$$

where $M(t) = E\{N(t)\}$ is the ordinary *renewal (counting) function*, given by:

$$M(t) = F(t) + \int_0^t M(t - x)dF(x) \ . \tag{6}$$

Many analytic results are known about $M(t)$, such as:

$$(t/m_1) - 1 \leq M(t) \leq F(t)/\bar{F}(t) \tag{7}$$

$$M(t) \approx f(0) \cdot t + [f'(0) + f^2(0)] \frac{t^2}{2} \qquad (t \to 0) \tag{8}$$

$$M(t) \approx \frac{t}{m_1} + \left(\frac{m_2}{2m_1^2} - 1\right) \qquad (t \to \infty) \tag{9}$$

with many more results known for specific F, or special shape assumptions, such as IFR, IFRA, NBU, etc.

However, here we merely note that (6) will only be $(m_1)^{-1}$ for all t in the special case of exponential lifetimes. In more general cases $M(t)/t$ begins at a value $f(0)$ (the intercept of the lifetime density) and can either grow or decline $\left[-f'(0) \lessgtr f^2(0)\right]$ for small values of t. For large values of t, $M(t)/t$ can either approach $(m_1)^{-1}$ from above or below depending upon whether the process is more or less "regular" $\left[m_2 \lessgtr \frac{m_1^2}{2}\right]$ than the exponential.

In short, the transient behavior of the mean value of $N(t)/t$ over a few average lifetimes can exhibit either apparent reliability "growth" or "decay," *even though the underlying process is stationary.*

If S systems were on test, and each reported N_i failures $(i = 1,2, \ldots, i)$, a *pooled estimator* would be:

$$\hat{\psi}(t) = \frac{\sum N_i(t)}{S \cdot t} \xrightarrow{a.s.} \frac{M(t)}{t} \tag{10}$$

as $S \to \infty$, giving the same result.

4. THE MEAN VALUE OF NAIVE MTBF ESTIMATOR

The average value of the naive MTBF estimator (2),

$$e_1(t) = E\{\mu_1(t)\} = E\{[Y_{N(t)}/N(t)] \mid N(t) > 0\} \tag{11}$$

has interesting properties. By direct argument

$$\Pr\{N(t) = n \,;\, Y_{n(t)} \in (y, y + dy]\} = f^{n*}(y)\bar{F}(t - y)dy \ , \quad y \in [0,t] \tag{12}$$

where f^{n*} is the n-fold convolution of f , so that:

$$e_1(t) = \frac{\sum_{n=1}^{\infty} \int_0^t \frac{y}{n} f^{n*}(y)\bar{F}(t - y)dy}{F(t)} \ . \tag{13}$$

The numerator can be simplified using LaPlace (-Stieltjes) transforms, giving the surprising results:

$$e_1(t) = \frac{\int_0^t yf(y)dy}{F(t)} = \int_0^t \left[1 - \frac{F(y)}{F(t)}\right] dy = E\{X \mid X < t\} \ , \tag{14}$$

the truncated mean of the underlying *distribution*.

From this, we observe:

(1) The average value of the naive estimator will *always* exhibit apparent "reliability growth" as t increases;

(2) Only if X has a finite range $[0,R]$ will $e_1(t)$ achieve its ultimate value, m_1 , for $t \geq R$.

More specifically, for small t ,

$$
(15) \quad \begin{aligned} e_1(t) &\approx \frac{t}{2} + \frac{1}{12}\left[\frac{f'(0)}{f(0)}\right] t^2 \; ; \; (f(0) \neq 0) \\ &\approx \frac{2}{3}\, t + \frac{1}{36}\left[\frac{f''(0)}{f'(0)}\right] t^2 \; ; \; (f(0) = 0) \quad (f'(0) \neq 0) \end{aligned}
$$

that is, the apparent growth in the estimator depends only on the shape of the density f near the origin, not on any specific values. For large t , the behavior of $e_1(t)$ depends upon the tails of F . Specifically, for algebraic tails, $\bar{F}(t) \approx A/t^{1+\alpha}$ for large t , $e_1(t) \approx m_1 - A(1 + \alpha)/(\alpha t^{\alpha})$.

The transient behavior of (14) can be quite long. For exponential lifetime, $\bar{F}(y) = e^{-\lambda y}$, we find:

$$
(16) \quad e_1(t) = \frac{1}{\lambda} \frac{[1 - (1 + \lambda t)e^{-\lambda t}]}{[1 - e^{-\lambda t}]} \; .
$$

From Figure 2 (solid line #1) we see that there is a significant bias in the naive MTBF for testing intervals less than, say, four MTBFs.

5. THE CLEVER MTBF ESTIMATE

The average of (3) cannot, in general, be put into simple form:

$$(17)\qquad e_2(t) = E\{\mu_2(t)\} = E\left\{\frac{t}{N(t)+1}\right\} = t\sum_{n=0}^{\infty}\frac{1}{n+1}P_n(t)\ ,$$

where $P_n(t)$ is the counting distribution associated with $N(t)$. However, since

$$(18)\qquad P_n(t) = F^{n*}(t) - F^{(n+1)*}(t)\ ,$$

where F^{n*} is the distribution corresponding to f^{n*} , we have the equivalent result

$$(19)\qquad e_2(t) = t - t\sum_{n=1}^{\infty}\frac{1}{n(n+1)}F^{n*}(t)\ .$$

For small t , we have the approximate result

$$(20)\qquad e_2(t) \approx t - \frac{f(0)}{2}t^2 - \left[\frac{f'(0)}{4} + \frac{f^2(0)}{12}t^3\right]\ .$$

Since $N(t) \to \infty$ and $\frac{N(t)}{t} \to 1/m_1$ with probability one as $t \to \infty$, $e_2(t)$ approaches m_1 . We suspect this growth is monotonic (apparent "reliability growth") in general, but have not been able to prove it. The exponential case can be summed explicitly, giving:

$$(21)\qquad e_2(t) = \frac{1}{\lambda}[1 - e^{-\lambda t}]\ .$$

Figure 2 (Solid line #2) shows that this estimator grows more rapidly than (16), but still shows significant bias. If we neglect the cases where $N(t) = 0$,

$$E\left\{\frac{t}{N(t)+1} \mid N(t) > 0\right\} = \frac{1}{\lambda}\,\frac{[1 - e^{-\lambda t}(1 + \lambda t)]}{1 - e^{-\lambda t}} \equiv e_1(t) \tag{22}$$

which shows in what sense the reliability engineer was clever about the incomplete sample.

6. THE SOPHISTICATED MTBF ESTIMATE

The average of the total-time-on-test estimator (4) is also difficult to simplify.

$$(23)\quad e_3(t) = E\{\mu_3(t)\} = E\left\{\frac{t}{N(t)} \mid N(t) > 0\right\} = \frac{t \sum_{n=1}^{\infty} \frac{1}{n} P_n(t)}{F(t)} .$$

Using (18), we obtain

$$(24)\quad e_3(t) = \frac{t\left[F(t) - \sum_{n=1}^{\infty} \frac{1}{n(n+1)} F^{n*}(t)\right]}{F(t)} .$$

For small (t), we obtain

$$(25)\quad \begin{aligned} e_3(t) &\approx t - f(0)\frac{t^2}{4} \qquad (f(0) \neq 0) \\ &\approx t - \frac{f''(0)}{f'(0)}\frac{t^2}{3} \qquad (f(0) = 0) \quad (f'(0) \neq 0) . \end{aligned}$$

Again, $e_3(t) \to m_1$ as $t \to \infty$, probably monotonically.

The exponential case can be expressed as:

$$(26)\quad e_3(t) = \frac{1}{\lambda}\,\frac{(\lambda t)^2 e^{-\lambda t} \sum_{j=1}^{\infty} \frac{(\lambda t)^{j-1}}{jj!}}{1 - e^{-\lambda t}} = \frac{1}{\lambda}\,\frac{\lambda t e^{-\lambda t} \int_0^{\lambda t} \frac{e^x - 1}{x}\,dx}{1 - e^{-\lambda t}} ,$$

and is shown in Figure 2, (Solid line #3).

Thus the mean of the sophisticated MTBF estimator grows more rapidly than the other two, but "over shoots" the true value by more than 30% at about four MTBFs! There it exhibits very slow "reliability decay," being still $1.056/\lambda$ at 20 MTBFs!

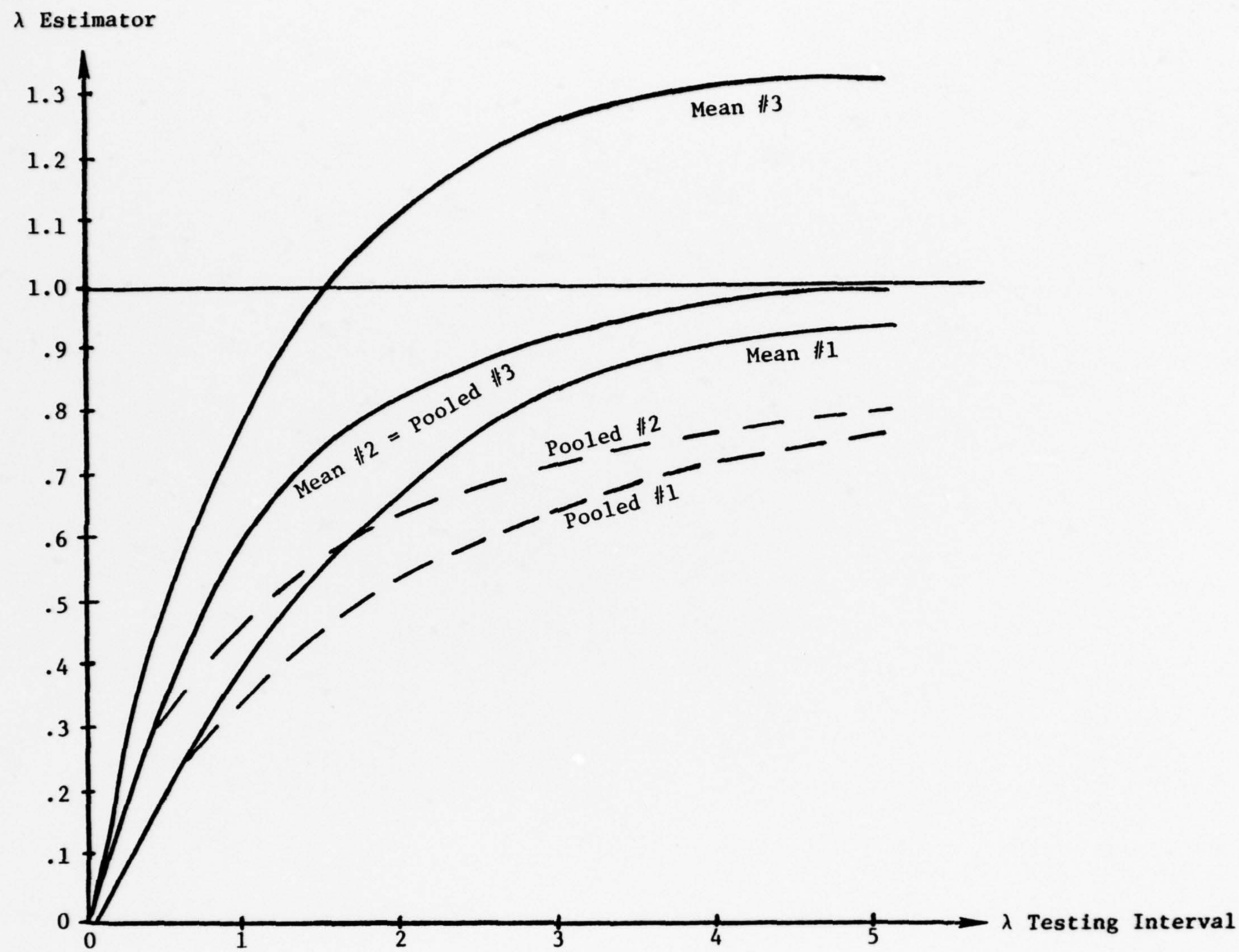

FIGURE 2: Growth of three estimators of λ^{-1} (MTBF) with testing interval for exponential samples. Solid line shows mean value of estimator with single system; dashed lines show limiting value of pooled tests of many systems.

7. POOLED ESTIMATORS

It might seem that pooling the estimators for many simultaneous systems on test might eliminate some of the above difficulties. We examine the limiting case when the number of systems is large.

For the naive estimator, assume that only $N_i(t) > 0$ and the corresponding epochs are reported from each system. Then

$$\text{(27)} \qquad \hat{\mu}_1(t) = \frac{\sum_{i=1}^{S} Y_{N_i}(t)}{\sum_{i=1}^{S} (N_i(t) > 0)} \xrightarrow{\text{a.s.}} \frac{E\{Y_{N(t)} \mid N(t) > 0\}}{E\{N(t) \mid N(t) > 0\}}$$

and it can be shown that in the exponential case:

$$\text{(28)} \qquad \hat{\mu}_1(t) = \frac{1}{\lambda}\left[1 - \frac{1 - e^{-\lambda t}}{\lambda t}\right].$$

From Figure (2) (Dashed line #1), we see that this is, in fact, *worse* than the mean of the unpooled estimator.

For the clever estimator, instead of averaging (3) over all systems, one would pool both the total interval and the number of (complete and incomplete samples) including cases where $N_i(t)$ is zero, so:

$$\text{(29)} \qquad \hat{\mu}_2(t) = \frac{St}{\sum_{i=1}^{S} (N_i(t) + 1)} = \frac{t}{\frac{1}{S}\sum N_i(t) + 1} \xrightarrow{\text{a.s.}} \frac{t}{M(t) + 1}$$

in the exponential case.

$$\text{(30)} \qquad \hat{\mu}_2(t) = \frac{1}{\lambda}\left[\frac{\lambda t}{\lambda t + 1}\right]$$

which, in Figure 2 (Dashed line #2) is also well below the mean value for a single system.

Finally, for the sophisticated estimator, who receives only data with at least one failure from each system, we can argue that, in the limit:

$$(31)\quad \hat{\mu}_3(t) = \frac{St}{\sum_{i=1}^{S} (N_i(t) > 0)} \xrightarrow{a.s.} \frac{t}{E\{N(t) \mid N(t) > 0\}} = \frac{tF(t)}{M(t)} .$$

In the exponential case

$$(32)\quad \hat{\mu}_3(t) \approx \frac{1}{\lambda} (1 - e^{-\lambda t})$$

which is identically $e_2(t)$ (21).

Although pooled estimates from several systems on test no doubt have greatly reduced variance, we must conclude that they, in fact, exhibit worse "reliability growth" bias than the corresponding mean estimators.

8. CONCLUDING REMARKS

In concluding, it is perhaps appropriate to mention some other observations about the reliability growth literature:

(1) There often seems to be confusion between the growth or decline of *hazard rate* function, $h(t) = f(t)/\bar{F}(t)$, and the underlying mechanisms (if any) of reliability improvement. The first operates in "local" time (since repair or replacement); the second operates in "global" time (since systems testing began). We have already seen how non-constant $h(t)$ can lead to growth or decline in the *average failure rate*, $M(t)/t$, with *no* underlying reliability growth mechanism.

(2) As pointed out in [2], if underlying reliability growth *does* exist [as in a time-varying Poisson failure process], then *successive intervals are not* independent, and, of course, not identically distributed. Thus, any methodology that assumes growth, but uses standard Poisson - or renewal - process methodology must be viewed with suspicion.

(3) Some literature speaks as if *total* operational time is significant - that is, a test over 5,000 for 1,000 hours each. In fact, we have seen that parallel testing may stabilize the estimator, but does not remove the transient bias over the real time testing interval.

We realize that more issues about reliability growth have been raised than resolved. Removal of testing biases in the stable case is only the first objective of research in this area. There is clearly a great need for additional work on extension of true, continuous reliability growth, as well as investigation of the statistical confidence of these estimators for different experimental protocols. We believe a correct formulation of this problem must use a Bayesian point of view, and incorporate the costs of experimentation into the design decision process. Only in this way will performance guarantee programs be placed on a sounder analytical basis.

REFERENCES

[1] Annual Reliability and Maintainability Symposium, 1978 Proceedings, Los Angeles, California, January 17-19, 1978, IEEE, N.Y.C.

[2] Ascher, H. and H. Feingold, "Is There Repair After Failure?," in [1], pp. 190-197.

[3] Balaban, H. S. and M. A. Meth, "Contractor Risk Associated with Reliability Improvement Warranty," in [1], pp. 123-129.

[4] Barlow, R. E. and R. A. Campo, "Total Time on Test Processes and Applications to Failure Data Analysis," in RELIABILITY AND FAULT TREE ANALYSIS, Barlow, Fussell and Singpurwalla, (eds.), SIAM, Philadelphia, (1975).

[5] Barlow, R. E. and E. M. Scheuer, "Reliability Growth During a Development Testing Program," Teknometrics, Vol. 8, pp. 53-60, (February 1966).

[6] Clarke, J. M. and W. P. Cougan, "RPM - A Recent Real Life History," in [1], pp. 279-285.

[7] Crow, L., "Reliability Growth Modelling," Technical Report No. 55, U. S. Army Material Systems Analysis Agency, Aberdeen (August 1972).

[8] Hauter, A. J. and C. W. Strempke, "Tacan RIW Program," in [1], pp. 62-65.

[9] Jewell, W. S., "Bayesian Life Testing Using the Total Q on Test," in THE THEORY AND APPLICATION OF RELIABILITY, WITH EMPHASIS ON BAYESIAN AND NONPARAMETER METHODS, Vol. 1, pp. 49-66, Academic Press, New York (1977).

[10] Lakner, A. A., R. T. Anderson and A. DiGianfilippo, "Cost Effective Reliability Testing," in [1], pp. 271-278.

[11] Lilius, W. A., "Reliability Growth Planning and Control," in [1], pp. 267-270.

[12] Newman, D. G. and L. D. Nesbitt, "USAF Experience with RIW," in [1], pp. 55-61.

[13] Shurman, M. B., "Time Dependent Rates for Jet Aircraft," in [1], pp. 198-203.

[14] Weiss, H. K., "Estimation of Reliability Growth in a Complex System with Poisson-Type Failure," Operations Research, Vol. 4, pp. 503-628, (October 1956).