CHARACTERIZING SYSTEM FAILURE CURVES WITH VITALITY

J. J. Anderson*
School of Aquatic & Fishery Sciences, University of Washington
Seattle, WA 98195

ABSTRACT

Identifying the causes of failure is fundamental to improving the survivability of Army systems whether they are remote sensors, complex vehicles, or networks. The shape of a repair frequency curve of a complex system is determined by variability in the initial quality of its components, variability in the system's wear and maintenance, and random failures independent of the condition of the components. The contributions of these sources of failure are modeled using vitality theory, which describes population survival curves in terms of the stochastic decline of a hidden Markov process (vitality) to an absorbing boundary representing death. By reinterpreting the vitality model, the population becomes the complex system, the individuals become the system's components, vitality becomes the remaining amount of component wear prior to failure, and absorption into the boundary represents component failure and replacement. Fitting the model to the repair frequency curve of a 180 ton rear dump truck, the shape of the repair curve is quantitatively partitioned into three factors: the operating environment, quality control in manufacture, and variability in maintenance. This information may be useful in identifying causes of system failure.

1. INTRODUCTION

Characterizing and quantifying the failure patterns of systems is a vast field that draws on theories from engineering, statistics and demography. Numerous models derived from these fields can successfully fit both failure data of repairable equipment and life-time survival data of nonrepairable equipment (e.g. Lieblein and Zelen, 1956; Cha and Mi, 2006). Well-known examples include mixed exponential functions (Vaupel and Yashin, 1985), non-homogenous Poisson processes (Pulcini, 2001) and the Weibull survival function (Mudholkar, Srivastava and Kollin, 1996). Perhaps the most commonly used models in the reliability analysis of repairable equipment are the non-homogeneous Poisson processes (Pulcini, 2001), which are based on the assumption that, when the equipment fails, the repair action returns the equipment to the condition it was in just before the failure occurrence (minimal repair). The above functions are largely empirical and so the rate parameters that characterize the curves often are not well defined in terms of the underlying processes that produce failure. For example, an exponential function simply assumes that the rate of failure is constant with age but it provides no reason why the rate should be constant or anything about the underlying mechanisms that produce failure. In contrast, survival/failure models based on an underlying hypothesis about the failure processes are desirable because the coefficients estimated by fitting the model to data can then be correlated with independent measures of the hypothesized process. If correlations between model coefficients and measures of the processes are statistically significant, we have some support of the model and quantification or ranking of the importance of the processes producing failure.

In this paper, I describe equipment failure patterns with a mechanistic model developed to characterize survival curves in biological populations. The model combines age-independent and age-dependent mortality processes into a single equation, the vitality equation (Anderson, 2000). The age-dependent source of mortality is described as the loss of vitality, a surrogate for remaining survival capacity of the organism. The stochastic rate of loss of vitality over the organism's life time is modeled with a hidden Markov process. This characterization of mortality is not new to biology or demography (Sacher, 1956; Strehler and Mildvan, 1960) and its mathematical characteristics are well established (Chhikara, R. S. and J. L. Folks, 1989; Aalen and Gjessing, 2001; Steinsaltz and Evans, 2004 and 2007). However, its applications to real issues in demography and ecology have been relatively limited (Weitz and Frazer, 2001; Anderson et al., 2008. However, because vitality-type models are developed from fundamental, but admittedly highly simplistic assumptions, they offer some new into the mechanisms of failure and mortality. In particular, an extension of the vitality model by Li (2008) offers a mathematically rigorous way of partitioning categories of failure processes that shape survival curves in organisms and failure curves of equipment. Here I explore the application of this extended vitality model to failure date of a repairable system, a 180 ton rear dump trunk, and propose that the model coefficients quantify, the failure processes into three categories: contributions of the operating environment, quality control in manufacture and variability in maintenance.

2. MODEL DEVELOPMENT

I first describe the vitality model as it applies to survival curves of biological populations and then discuss how the assumptions of the model are reinterpreted to describe the failure curves of repairable systems.

2.1 Vitality theory

Using the notation in Anderson (2000), consider the accumulation of damage leading to organism mortality in terms of the vitality, an abstract stochastic measure of the remaining survival capacity of the individual. Each individual begins with an initial vitality, v_0, and dies when its vitality reaches zero (Fig. 1). The rate loss of vitality is described by a continuous Markov process that is expressed as a stochastic differential equation with mean, ρ, and variation, σ, in the rate of loss of vitality

$$\frac{dv}{dt} = -\rho + \sigma\xi(t) \qquad (1)$$

where ξ is a rapidly fluctuating random term or white noise with a zero mean and unit intensity. The integral of the white noise process is typically assumed to have a Gaussian distribution (Gardiner, 1985).
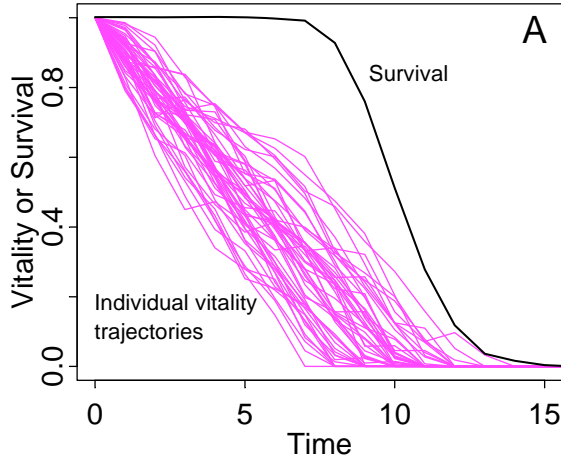

Fig. 1. Individual vitality trajectories defined by Eq. (1) and the resulting survival described by Eq. (4) with $u = 0$.

If all individuals have the same initial vitality, v_0, then the vitality probability distribution in the population remaining at time, t, is

$$p_v(t) = \frac{1}{\sigma\sqrt{2\pi t}}\left[\begin{array}{l} \exp\left(-\frac{(v - v_0 + \rho t)^2}{2t\sigma^2}\right) - \\ \exp\left(\frac{2\rho v_0}{\sigma^2} - \frac{(v + v_0 + \rho t)^2}{2t\sigma^2}\right) \end{array}\right] \qquad (2)$$

Fig. 2 illustrates the time varying change in of the density of vitality in the population according to Eq. (2). The population starts with an initial vitality, $v_0 = 1$, and evolves into a quasi-steady distribution in 12 time units. The area under the curve decreases with time as the density is absorbed into the zero boundary, i.e. a process representing death. Thus, because Eq. (2) is based on the initial distribution being Dirac delta function of unit area, the area under the curve at any $t > 0$ is equivalent to the fraction of the remaining population that has not died from the loss of vitality.
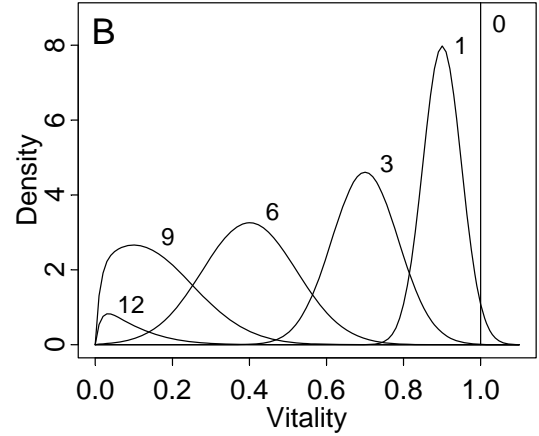

Fig. 2. The vitality density defined by Eq. (2) evolves from a Dirac distribution at time 0 into a Gaussian distribution over times 2 through 6 and into a gamma-like distribution by time 9. The area under the curve diminishes by the absorption of vitality density into the zero boundary.

Integrating Eq. (2) over the allowable range of vitality $(0, \infty)$ yields the fraction of the population at t that has not died of vitality related causes. This is shown as the survival curve in Fig. 1. Mortality can also occur from processes not associated with vitality, i.e. accidental mortality, and in the vitality model this is described with a Poisson mortality process with rate k. The details of this three-parameter model are given in Anderson (2000) and Anderson et al. (2008). An algorithm to estimate model parameters (r, s, k) from survival data is given in Salinger et al. (2006).

The vitality model based on Eq. (2) assumes all individuals are initially identical, which is in fact violated to a greater or lesser degree by every real population. To correct this weakness, Li (2008) extended the model by expressing the initial vitality as a Gaussian distribution

$$p(v_0) = \frac{1}{\sqrt{2\pi\tau}}\exp\left(-\frac{1}{2}\left(\frac{v_0 - \mu}{\tau}\right)^2\right) \qquad (3)$$

2

where μ and τ are the mean and standard deviation of the initial vitality distribution. The resulting equation, expressed in terms of the cumulative mortality, which is the appropriate measure for failure curves, is

$$f(t) = 1 - \left(\Phi\left(\frac{1 - rt}{\sqrt{u^2 + s^2 t}} \right) - \exp\left(\frac{2u^2 r^2}{s^4} + \frac{2r}{s^2} \right) \Phi\left(\frac{\frac{2u^2 r}{s^2} + 1 + rt}{\sqrt{u^2 + s^2 t}} \right) \right) e^{-kt} \quad (4)$$

where Φ is the normal cumulative distribution and the coefficients are

$$r = \rho/\mu, \quad s = \sigma/\mu, \quad u = \tau/\mu. \quad (5)$$

where r and s are the normalized mean and variability in the rate of loss of vitality, with dimension of $1/time$ and $1/\sqrt{time}$ respectively, u is the coefficient of variation in the initial vitality distribution and k the rate of accidental mortality with a dimension of $1/time$.

2.2 Estimating coefficients

The coefficients in Eq. (4) can be estimated with maximum likelihood techniques developed by Salinger et al. (2003) and modified by Li (2008) with a simulated annealing search algorithm. The details of the algorithm and its code written in the R® statistical language will appear in a forthcoming publication by Li and Anderson. The algorithm estimates the parameters, their standard errors and p-values.

2.3 Applying vitality to repairable systems

To apply the vitality model described in Eq. (4) to failure data of a repairable system, such as a vehicle, the system is taken as equivalent to the biological population and its components are equivalent to the organisms in the population. Organism vitality is reinterpreted as the remaining age-dependent wear of the components. The absorbing boundary at $v = 0$, which in a biological population represents the removal of individuals by death, becomes the removal of worn components by replacement and repair. In this construct, the model tracks the wear of the original components that have yet to fail. When an original component fails it is replaced with a component that has less wear, i.e. it is in better condition than the original part. Thus, the model assumes that as the system ages the number of components that can fail because of wear declines as the old parts are replaced with new or reconditioned ones. With these reinterpretations, the distribution of remaining wear the components can tolerate prior to failure is mathematically equivalent to the

distribution of vitality in a population. The model also contains an accidental failure rate, which is independent of the state of wear in the components. Thus, the model describes failure curves of repairable equipment in terms of age-dependent and age-independent processes.

2.4 Attributing meaning to the coefficients

Considering first the age-dependent process, we represent the average rate of wear by r, which is the average rate of loss of vitality in a biological population. We represent the variability in the wear rate of components by s. This term represents an evolving heterogeneity in the rate of wear in that as the system ages the distribution of wear amongst the components increases. One possible cause of this heterogeneity could involve routine service and maintenance. For example, it seems plausible that variations in cleaning and operating within design specification could increase the variation in the wear rates of components.

The model implicitly assumes that replacement components do not fail over the usable lifetime of the system. However, in reality they can which would alter the estimates of the model coefficients. Resolving effect of this assumption violation is beyond the scope of this paper, but it could readily be evaluated with a numerical analysis of the model.

A second major component of wear involves the initial quality of the system's components. In an ideally designed system, the expected lifetimes of all components are equal. This equality is desirable so the system does not contain expensive over-designed components that outlast the useful life of the system, or correspondingly, under-designed components that fail early necessitating costly repairs and recalls. In reality, the components have individual lifetimes that are distributed about the system's mean useful lifetime. The vitality model represents this variability as a distribution of initial component quality with coefficient of variation u. It is distinguished from the evolving heterogeneity noted above, in that the heterogeneity in component wear attributable to u does not change with system age while the heterogeneity attributable to different wear rates does change with age.

The final process we need to consider is the age-independent failure rate characterized by k in Eq. (4). By definition, this factor is independent of the age of a component and occurs at a constant random rate over the usable life of the system. Although the causes of random age-independent failure are difficult to identify, we expect in part they are related to the state of the operating environment of the system. For example, a collision of one vehicle with another falls readily into this category since the event is independent of the age of either vehicle,

but the probability of occurrence strongly depends on the environment in which the vehicles operate.

2.5 Break-in period and the left-truncation effect

Failure rate date of repairable equipment often exhibit a bathtub-type profile in which the rate is high during a burn-in or break-in period, then declines over the useful life of the system and again increases as the useful life is exceeded. The break-in period is common in manufactured systems and results from manufacturing error, defective parts and assembling defects. In repairable systems, these defects are identified and the break-in period is generally identified by a clear discontinuity in the failure rate. The vitality model does not capture this discontinuity, but the effect can be assessed by fitting the model to the data with and without the inclusion of the break-in period. However, by left-truncating data to remove the break-in period, we affect the estimate of u, because the evolving heterogeneity that occurs in the break-in period then becomes included in the initial heterogeneity of the left-truncated data fit.

The effect of excluding the break-in period data can readily be evaluated by computing estimates of model coefficient with different left-truncation intervals. The resulting coefficients must then be adjusted for the differences in interval length by the formulas

$$ r_i \approx r_j \frac{T_{\max} - T_i}{T_{\max} - T_j} \quad \text{and} \quad s_i \approx s_j \sqrt{\frac{T_{\max} - T_i}{T_{\max} - T_j}} \qquad (6) $$

where i and j are indices for break-in periods, T_i and T_j and $T_{\max}$ is the total length of the failure record.

2.6 Issues with the number of components

In biological survival data, the initial population size is known and the survival curve over time is well defined. However, in failure data, the population size is not necessarily well defined. The total number of failures, N, may involve multiple failures and repairs of individual components, repairs without replacement of parts, or replacement of parts with new or reconditioned equivalents. When fitting the model to the failure curve of a single system, N is simply the total number of repairs over the usable lifetime of the system and the coefficients estimated are specific to the categories included in the definition of repair. If the failure curves of a group of systems are being compared then the definition of failure must standardized and N should be set by the member of the group that experienced the most failures over the usable lifetime of all the systems. While, N will affect the parameter estimates, it is unlikely to radically alter the relative significance of the parameters. However, this extended analysis is beyond the scope of the paper.

3. EXAMPLE: DUMP TRUCK FAILURE

3.1 Data

Here we consider data on 128 failure times of a 180-ton rear dump truck (Coetzee, 1996, extracted from Pulcini, 2001). The truck was composed of multiple subsystems that could each independently fail resulting in a repair event.

3.2 Results

Using Li's (2008) simulated annealing algorithm the model coefficients were estimated by excluding the break-in period with three different left-truncation intervals T (Table 1). All break-in periods yielded similar estimates of r when adjusted according to Eq. (6). The coefficients normalized to no break-in period, $T = 0$, become $r = 0.476 \pm 0.014$ (yr^{-1}) and $s = 0.004 \pm 0.003$ (yr$^{-1/2}$). The mean wear rate estimate has a low standard error. Further, the mean lifetime is approximately $1/r$, which gives 2.05 to 2.16 operating years (17958 to 18912 operating hrs). In comparison, the wear-rate variability, s, is small and contains larger uncertainty. The accidental failure rate, k, is not affected by the truncation and the three estimates cluster within the standard error. The coefficient of variation of the component's initial quality, which is quantified by u, has a low standard error and increases with increasing truncation period as is expected by the theory.

Table 1. Coefficient values from Eq.(4). Standard errors (se) and left-truncation beak-in period T.

T (days)		r (yr^{-1})	s (yr$^{-1/2}$)	k (yr^{-1})	u
0	value	0.464	0.0050	0.233	0.117
	se	0.007	0.0417	0.038	0.006
1.7	value	0.479	0.0006	0.208	0.122
	se	0.007	0.0207	0.0369	0.058
4.9	value	0.513	0.0062	0.211	0.133
	se	0.009	0.0425	0.039	0.004

Failure curves for two left-truncation periods are shown in Fig. 3 and Fig. 4. The model fits the data well in all cases.
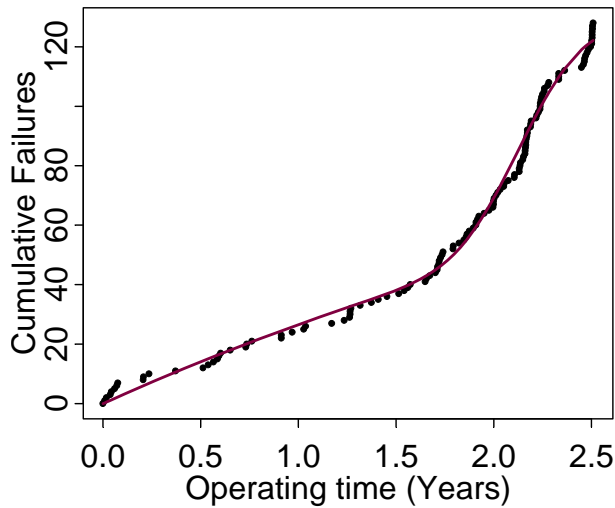

Fig. 3. Observed failure time of a rear dump truck (●) and model fit (—) with no left-truncation ($T = 0$). Data from Coetzee (1996) extracted from Pulcini (2001).
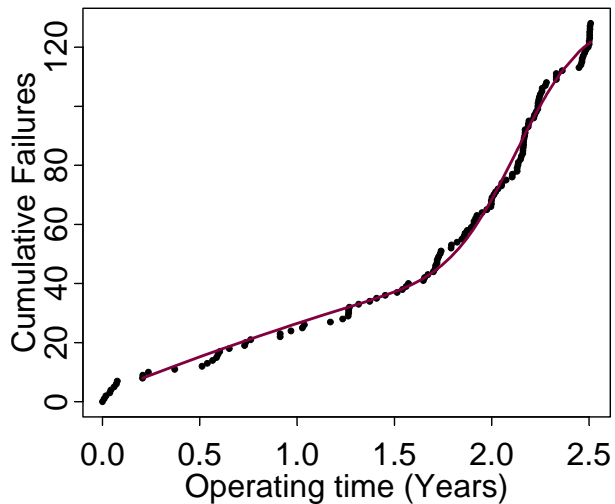

Fig. 4. Model fit with a left-truncation of 4.9 days.

3.3 Identifying factors of failure

The vitality equation attributes failure to random events that are independent of age and quantified by a constant rate (k), and age-dependent events that depend on an average wear rate (r), an initial heterogeneity in the quality of the system components (u) and an evolving heterogeneity in the wear rate (s). The relative effects of these factors on system failure curves are illustrated in Fig. 5 for estimates in Table 1 with $T = 0$. The individual sub-factor failure curves have several notable features.

The accidental failure rate (Fig. 5, Curve A), is generated with $f(t) = 1 - \exp(kt)$ and accounts for about 40% of the failures over the 2.5 yr operating-time of the truck. Curve B, generated by setting $u = k = 0$ in Eq.(4), characterizes the shape of the failure curve from evolving heterogeneity in the wear rate. The step-like failure pattern results because with all components having the same initial quality and with little variability in wear rates all components fail at essentially the same time. Thus, we may hypothesize that variance in routine maintenance is small and contributes little to the shape of the failure curve. However, the location of the step, ~2.1 yrs, is a measure of the average lifetime of the truck's components. Curve C, generated by setting $s = k = 0$ in Eq. (4), depicts the effects of variations in the initial quality of the truck's components on the failure curve. The variability in initial quality accounts for ± 0.5 yr variation in failure time. Correspondingly, the weakest parts of the truck have a useable life of 1.5 yrs. Finally, at the cross point of the curves (~2.1 yrs) the cumulative number of failures from accidental and wear-dependent processes are equal. However, at this age the rate of failure from wear-dependent processes is vastly larger that the rate from accidental failure and failure after the cross point is dominated by wear.
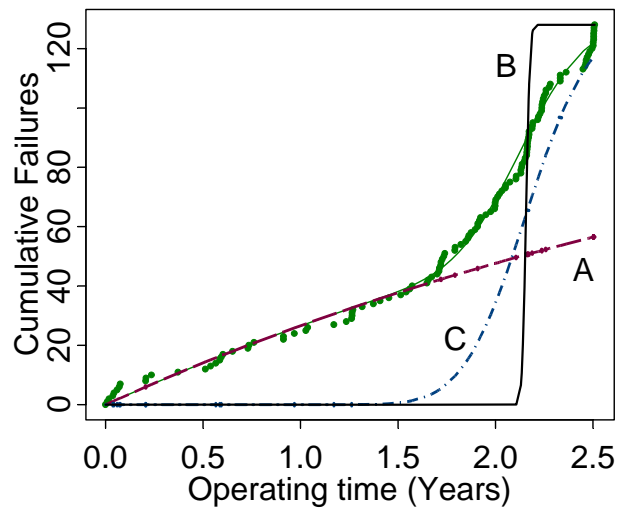

Fig. 5. Model fit to failure data (●) and curves attributed to the different failure components. A) component failure curve with random events only, B) failure curve with evolving heterogeneity in age dependent failure, C) failure curve with the intrinsic heterogeneity in failure components.

CONCLUSIONS

The key scientific contribution of the model presented in this paper is its mathematically rigorous, but simple, representation of failure and repair of complex systems through an analogy to biological vitality theory. This foundation contributes to the model's ability to fit repair curves and provides realistic meanings to the coefficients, which are the mean and variance in the wear rate, a distribution of initial component quality and a failure rate independent of wear. This partitioning of factors of failure into intuitive and meaningful categories is unique. In particular the possibility of identifying the independent effects of quality control in manufacturing and maintenance on failure is not found in traditional failure models.

The model should have value in improving both the manufacture and maintenance of complex equipment used by warfighters. As an example, fitting the model to repair records of LMTV A1 Cargo vehicles would generate distributions of the four model parameters for the population of vehicles. Then an analysis of the parameter distributions against factors hypothesized to affect vehicle failures could potentially identify significant factors in vehicle failures. For example, a significant relationship between the quality measure, $u,$ and manufacturing site might be used to rank the quality control of the different sites. Furthermore, under the model's assumptions and interpretations this site quality ranking would be independent of factors related to field maintenance and the vehicles' operating environments. Field maintenance and operating environments could be ranked in a similar manner and depending on the statistical significance of such analyses it may be possible to identify substandard steps in the vehicles' usable life cycle and thus direct corrective actions to specific issues.

In summary, the aim of this paper is to describe vitality-based failure modeling and its potential for quantitatively assessing the quality, maintenance and wear patterns of complex Army systems.

REFERENCES

Aalen, O. O. and H. K. Gjessing, 2001: Understanding the shape of the hazard rate: a process point of view. *Statist. Sci.*, **16**, 1–22.

Anderson, J. J., 2000: A vitality-based model relating stressors and environmental properties to organism survival. *Ecol. Monographs*, **70**, 445–470.

Anderson, J. J., M. C. Gildea, D. W. Williams and T. Li, 2008: Linking growth, survival and heterogeneity through stochastic vitality. *Am. Nat.*, 171, No.1 E-article.

Cha, J. H. and J. Mi, 2007: Some probability functions in reliability and their applications. *Nav. Res. Logist.*, **54**, 128–135.

Chhikara, R. S. and J. L. Folks, 1989: The inverse Gaussian distribution: theory, methodology, and applications. New York: Marcel Dekker.

Coetzee, J. L., 1996: Reliability degradation and the equipment replacement problem. *In: Proc. Int. Conf. of Maintenance Societies (ICOMS-96)*. Melbourne, Australia, 1996, Paper 21.

Gardiner, C. W., 1985: Handbook of stochastic methods for physics, chemistry, and the natural sciences. Springer-Verlag, Berlin.

Li, T., 2008: The Extension of the Vitality Model and Its Application. Master of Science thesis, University of Washington, Seattle.

Lieblein, J. and M. Zelen, 1956: Statistical investigations of the fatigue life of deep-groove ball bearings. *J. Research Nat'l. Bur. Stds.*, **57**, 273–316.

Mudholkar, G. S., D. K. Srivastava and G. D. Kollin, 1996: A generalization of the Weibull distribution with application to the analysis of survival data. *J. Am. Statist. Assoc.*, **91**, 1575–1583.

Pulcini, G., 2001: Modeling the failure data of repairable equipment with bathtub type failure intensity. *Reliab. Eng. Syst. Saf.*, **71**, 209–218.

Sacher, G. A., 1956: On the statistical nature of mortality with special reference to chronic radiation mortality. *Radiology*, **76**, 250–257.

Salinger, D. H., J. J. Anderson, and O. S. Hamel, 2003: A parameter estimation routine for the vitality-based survival model. *Ecol. Modeling*, **166**, 287–294.

Steinsaltz, D., and S. N. Evans, 2004: Markov mortality models: implications of quasistationarity and varying initial distributions. *Theor. Popul. Biol.*, **65**, 319–337.

Steinsaltz, D., and S. N. Evans, 2007: Quasistationary distributions for one-dimensional diffusions with killing. *Trans. Amer. Math. Soc.*, **359**, 1285–1324.

Strehler, B. L., and A. S. Mildvan, 1960: General theory of mortality and ageing (a stochastic model relates observations on aging, physiologic decline, mortality, and radiation). *Science*, **132**, 14–19.

Vaupel, J. W. and A. I. Yashin, 1985: Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics. *Amer. Statistician*, **39**, 176–185.

Weitz, J. S., and H. B. Fraser, 2001: Explaining mortality rate plateaus. *Proc. Natl. Acad. Sci. USA* 98,15383–15386.

SECURITY CLASSIFICATION & PUBLIC INFORMATION RELEASE FORM
26th Army Science Conference

The purpose of this form is to document the security, OPSEC, and public information reviews of papers submitted for the 26th Army Science Conference (26th ASC) and to authorize presentation, videotaping, audiotaping and publication to CD/web (www) of papers selected for presentation at the 26th ASC, JW Marriott Orlando Grande Lakes, Orlando, FL, 1-4 December 2008. This form must accompany the final summary and manuscript. Without its completion, the summary and manuscript will not be considered for presentation or publication. Please complete all blocks and affix signatures.

TITLE OF PAPER: _Characterizing System Failure Curves With Vitality_

NAME AND ORGANIZATION OF PRINCIPAL AUTHOR: _J. J. Anderson, University of Washington, School of Aquatic and Fishery Sciences_

TELEPHONE: (Commercial Only): _206-543-4772_

PART I. SECURITY/OPSEC REVIEW - The final summary and manuscript must be reviewed by an authorized Security Officer IAW guidance under AR 380-5, DoD Directive 5200.20, AR 530-1, the Military Technologies List, and any other applicable security regulations. The paper must be unclassified and may not contain Restricted Data, Formerly Restricted Data or be marked NOFORN.

1. [✓] The summary is **UNCLASSIFIED/ PUBLIC DOMAIN.** (Only unclassified/ public domain summaries will be accepted.)

2. [✓] The final manuscript is **UNCLASSIFIED/ PUBLIC DOMAIN** and suitable for release and publication.

3. [✓] The manuscript and summary have undergone OPSEC review, and it has been determined that the material is suitable for release and publication in the public domain.

4. [✓] The proposed presentation has been reviewed to assure that classified information is not disclosed and that material is suitable for release to DoD attendees and cleared representatives of NATO/ ABCA countries.

5. Typed Name, Title, Grade, and Organization of OPSEC Reviewer   Signature of OPSEC Reviewer

TIM RAINES, ERDC OPSEC Officer, GG-11 _for Linda D. McGowan_ _23/ May_ 2008

6. Typed Name, Title, Grade, and Organization of Security Reviewer   Signature of Security Reviewer

ERDC Command
LINDA D. McGOWAN, Security Manager, GG-13 _Linda D. McGowan_ _23 / May_ 2008

PART II. INFORMATION RELEASE REVIEW. All papers must be reviewed for suitability of release (IAW AR 380-5), editorial accuracy, content and propriety. The reviewer must be knowledgeable of the research presented in the paper and cannot be an author or coauthor of said paper. No additional editing of this manuscript will be made prior to publication.

7. [✓] The summary and manuscript have been reviewed for conformance with release policy, editorial accuracy, content and propriety. They are suitable for presentation and publication in the public domain.

8. Typed Name, Title, Grade, and Organization of Information Release Reviewer   Signature of Information Release Reviewer

RICHARD E. PRICE, PhD, Chief, Environmental
Processes & Engrg Division, DB-5, ERDC   _[signature]_ _5 / 23/_ 2008

Please fax completed form to 757-357-5108.