9-20-1999

Statistical Tools for the Rapid Development & Evaluation of High-Reliability Products

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Michael Hamada
University of Waterloo

Statistical Tools for the Rapid Development & Evaluation of High-Reliability Products

William Q. Meeker
Iowa State University, Ames

Michael Hamada
University of Waterloo, Waterloo

September 20, 1999

Abstract

Today's manufacturers face increasingly intense global competition. To remain profitable, they are challenged to design, develop, test, and manufacture high reliability products in ever-shorter product-cycle times and, at the same time, remain within stringent cost constraints. Design, manufacturing, and reliability engineers have developed an impressive array of tools for producing reliable products. These tools will continue to be important. However, due to changes in way that new product-concepts are being developed and brought to market, there is need for changes in methods used for design-for-reliability and reliability testing, assessment, and improvement programs.

This paper uses a conceptual degradation-based reliability model to describe the role and need for integration of reliability data sources. These sources include accelerated degradation testing, accelerated life testing (for materials and components), accelerated multifactor robust-design experiments and over-stress prototype testing (for subsystems and systems), and the use of field data (especially early-production) to produce a robust, high-reliability product and to provide a process for continuing improvement of reliability of existing and future products. Manufacturers need develop economical and timely methods of obtaining, at each step of the product design and development process, the information needed to meet overall reliability goals. We emphasize the need for intensive and effective upstream testing of product materials, components, and design concepts.

Key Words: Accelerated testing; Accelerated life testing; Degradation measurements; Designed experiments; Prototype testing; Robust-design; Field data feedback; Failure mode analysis.

1 Introduction

1.1 Quality and reliability

Rapid advances in technology, development of highly sophisticated products, intense global competition, and increasing customer expectations have combined to put new pressures on manufacturers to produce high quality products. Customers expect purchased products to be reliable and safe. Systems, vehicles, machines, devices and so on should, with high probability, be able to perform their intended function under encountered operating conditions, for some specified period of time.

Reliability is often defined as the probability that a system, vehicle, machine, device, and so on, will perform its intended function under encountered operating conditions, for a specified period of time. Improving reliability is an important part of the larger overall picture of improving product quality. There are many definitions of quality, but general agreement that an unreliable product is *not* a high quality product. Condra [4] emphasizes that "reliability is quality over time." In reliability there are strong parallels with (or generalizations of) important quality concepts such as metrics, process capability, and continuous improvement.

Statistical process monitoring and designed experiments have become important tools for design and process engineers in their quest for continuous quality improvement. Methods of improving quality have had a positive impact on improving reliability. Using methods with particular focus on reliability, however, has the potential to produce additional important improvements that would not be possible by using quality improvement methods alone.

1.2 Difficulty in measuring reliability

Modern programs for improving reliability require quantitative methods for predicting and assessing various aspects of product reliability. Careful planning is required for efficient and timely collection of the needed data. This will involve the collection of "reliability data" from combinations of studies such as laboratory life tests, degradation tests of materials, devices, and components, designed experiments for reliability improvement, tests on early prototype units to learn about possible failure modes, careful monitoring of early-production units in the field, analysis of warranty data, and systematic longer-term tracking of product in the field.

Among quality metrics, reliability is one of the more difficult to monitor and control; it can be assessed *directly* only after a product has been in the field for some time. Similarly, identification and rectification of root causes of reliability problems can be more difficult. Predictions or *indirect* measures of reliability based on laboratory testing have the potential to be highly inaccurate.

1.3 Risks and needs for using new (unproven) technology

The difficulty of the challenge to improve product reliability is compounded by tighter regulations on, for example, fuel economy, auto emissions, and the release of harmful chemical compounds into the environment. These regulations are requiring the introduction of materials and technologies that have not, as of yet, met the test of time. In some products, large safety factors of past designs are being reduced to meet weight and cost constraints. Changes in existing technology, materials, and safety factors dictate the need for increased amounts of testing, development, and experimentation.

1.4 Need for rapid product development

In opposition to the need for more testing to achieve high quality and reliability, there is, today, extreme pressure to reduce the amount of time needed to go from new-product conceptualization to production. This pressure has the effect of further reducing the amount of time available for reliability testing at all stages of the product-development process. This creates new challenges for product and production process design and for planning and analyzing the results of reliability studies to reduce the risk of overlooking potential future reliability problems.

1.5 Overview: types of failures and sources of information for reliability improvement

This paper is organized as follows. Section 2 contrasts traditional ideas of reliability demonstration with more modern concepts of reliability assurance. Section 3 describes a simple classification of failure types. Section 4 describes a generic model for failures, providing a framework for our subsequent discussion of reliability tests and other sources of information for reliability improvement/assurance programs. Section 5 describes the important differences between traditional time-to-failure data and the more generally useful degradation data.

Table 1 outlines our subsequent discussion of different types of reliability tests and the types of failures they are designed to find. Depending on the particular industry and product being considered, there might be some disagreement with the placement of the X's in the table, but the general idea is that different types of reliability tests are needed at different stages of product development to find and obtain information about different kinds of failure modes. In general, upstream tests (Section 6) will characterize and qualify components and materials as well as provide information on degradation rates and/or time-to-failure distribution of failure modes known occur in actual field operation. Most product design changes made early in the product-development cycle (upstream) are relatively painless and inexpensive (i.e., relative to changes made after a product introduction to customers). An important class of upstream reliability tests (Section 7) focuses on discovering unexpected failure modes and improving reliability involving interactions and interfaces among product components and subsystems. Downstream reliability tests (Section 8) are used to verify or demonstrate final product reliability or, in some cases, used as a screen to remove defective product before shipping. There can be serious economic and other consequences when a failure mode is discovered in downstream tests or in actual service. Section 9 describes the important role of field data in the reliability improvement/assurance process. Section 10 summarizes some of the important ideas and explains some of the key ideas behind information flows in the design of an effective reliability improvement/assurance process.

Table 1: Types of failure found by different reliability tests

		Kind of Failure Mode				
	Type of Test	Unavoidable Degradation Failure	Known Infant Mortality	Unexpected Infant Mortality	Known Accidental Failure	Unexpected Accidental Failure
Upstream	ALT	X	X			
	ADT	X	X			
	Static RDE's	X	X	X	X	
	Dynamic RDE's	X	X	X	X	
	STRIFE Tests		X	X		
Downstream	ESS/Burn-in Screening Tests			X	X	
	Early Field Data		X	X	X	X
	Other Field Data	X	X	X	X	X

1.6 Acronyms

This paper uses the following acronyms.

ALT	Accelerated Life Test	FMEA	Failure Modes and Effects Analysis
ADT	Accelerated Degradation Test	RDE	Robust-design experiment
ESS	Environmental Stress Screening	STRIFE	Stress-Life Test

2 Reliability Demonstration and Reliability Assurance Processes

This section contrasts some of the traditional ideas of reliability demonstration with more modern ideas of developing a process for reliability assurance.

2.1 Reliability demonstration

Manufacturers may be expected to "demonstrate" to customers (internal or external) that their products have a specified degree of reliability. For high reliability products, with limitations on the number of samples and amount of available testing time, such demonstrations are expensive, impracticable or impossible. Demonstration puts the burden of proof on the manufacturer; an inconclusive test results in a failed demonstration. For example, to demonstrate, at a 95% level of confidence, that a system will operate for five years with reliability 0.999 would require that 2995 units be tested for five years without failure.

To overcome the difficulties of conducting timely and economical reliability demonstration tests, a number of approaches have been suggested and used, with varying degrees of success.

- For some products, testing time can be compressed by a) increasing use-rates and/or b) using stress-acceleration factors (e.g., temperature, humidity, and voltage) to make units fail faster. For complicated products it is difficult or impossible to achieve much in the way of time acceleration in end-of-production demonstration tests. This is because final products typically have many subsystems and other components. In particular

  - Products may have many potential failure modes, each affected differently by the stress-acceleration factors (raising the possibility of masking of practically important failure modes by those that would be more highly accelerated).

  - Some system components will have limited ability to tolerate increased stresses without incurring damage, causing failures that would not be seen in actual application (e.g., melting of plastic).

  - *Extrapolation in stress* is extremely dangerous outside of situations where there is only one or a small number of failure modes for which the kinetics of failure are well understood and/or extrapolation models have been verified empirically.

  It is for these reasons that accelerated testing is most useful for obtaining component-level information.

- With the introduction of assumptions in the form of a particular time-to-failure distribution (e.g., the Weibull distribution), it may be possible to decrease the length of the test and maintain precision. Then reliability predictions require *extrapolation in time* (e.g., using data over 6 months to predict out 5 years). Needed sample sizes remain high and there is potential for having serious model-extrapolation errors (e.g., the human mortality extrapolation example in [7]).

- Use of Bayesian methods for demonstration testing allows the incorporation of "prior information" into the demonstration. With strong prior information it is possible to obtain precise results with smaller sample sizes. The danger is that the prior information may reflect wishful thinking instead of accurate reliability information, leading to unrealistically optimistic reliability predictions, as described in [6]. Even if the prior information has basis in past experience, its use requires a third form of dangerous extrapolation—*extrapolation from past product.*

In addition to the other assumptions listed in this section, reliability demonstration tests have an implicit assumption that the test conditions reflect actual use conditions. See [8] for detailed discussion of assumptions underlying statistical analyses from enumerative and analytic studies.

2.2 Traditional reactive approach taken when reliability requirements are not met

In spite of best intentions and efforts, in many cases reliability goals and requirements are not met at or near to the time product is to be delivered to a customer or customers. In the worst cases, the reliability problems become known after large numbers of units have been introduced into the field. At this stage, various alternatives need to be considered (or pre-planned contingencies implemented). Some of the alternatives (depending on the particular reliability problem and economic considerations) include

- Delay product introduction and continue (or begin) a program of reliability growth involving a test-and-fix process (design changes to improve reliability).
- Use burn-in or environmental stress screening to eliminate defective components from systems before they are shipped.
- Product recall or field changes of critical components.
- Agreements to cover losses through warranty protection to cover expenses due to field product failures.

Any of these will lead to some combination of additional expense, increased inconvenience, loss of revenue and goodwill, delay of product introduction, uncertainty about future product reliability, and delay of efforts in the next generation of product development.

2.3 Modern proactive approach of reliability assurance—the roles of reliability tests, experiments, and use of past field data

The development of *high-reliability* products in shorter product-cycle times will require the use of *reliability assurance processes*. Figure 1 provides an example of such a process. The solid boxes and lines show the core activities in design production and product flow. The dashed boxes and lines show statistical and engineering information sources and information flow paths. Generally, reliability assurance processes should concentrate most reliability testing efforts upstream in the product conceptualization/design/development sequence. This is analogous to using upstream process design and process monitoring instead of downstream inspection to produce high-quality products. In the rest of this paper we will discuss a unifying model and classification for product failure and discuss the statistical tools outlined in Figure 1.

3 A Simple Classification of Failure Types

Understanding the physical and chemical mechanisms or random risk factors leading to failure can suggest effective methods for eliminating failure modes or reducing the probability of a failure mode. Broadly speaking, it is possible to classify failure modes into three different categories which we describe in this section.

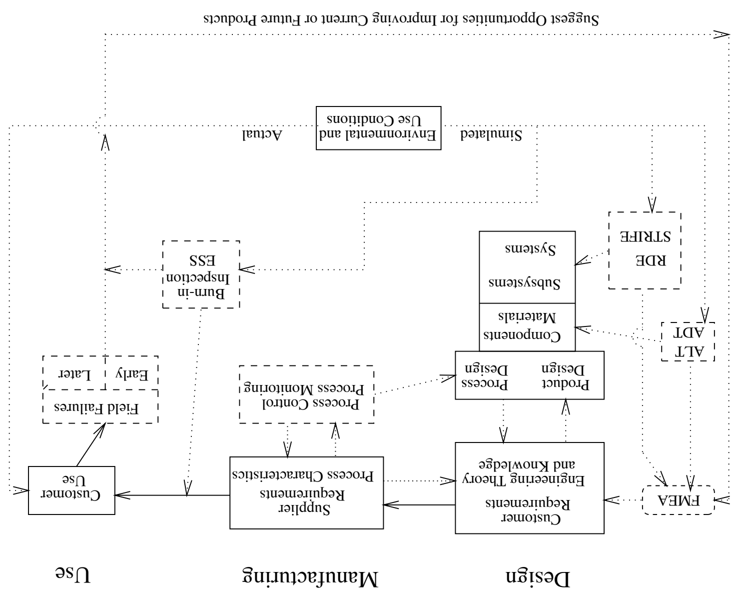

Figure 1: Example of information flow patterns in a reliability assurance process

3.1 Infant mortality failures

Infant mortality failures have been most common in electronic equipment, but are possible in most any kind of product, especially those using new or otherwise untested technology. Infant mortality results from units that contain manufacturing defects that will lead to premature failure. Infant mortality failures are often traced to a defect-causing degradation mechanism. Examples include

- A crack in the encapsulement of an integrated circuit allowing destructive reactive elements to enter.
- Improper installation of a mechanical component causing higher than usual wear rates.
- Improper preparation of a surface before painting.

Often such defects tend to concentrate in particular batches. Generally the proportion of such defects is small and will decrease as a product design and manufacturing process matures.

For infant mortality failure modes, the primary concern is not the time-to-failure distribution but instead with the proportion of units that will fail unexpectedly. See [18] for a description of statistical methods for analyzing infant mortality data. For short-term economic reasons, failures during the warranty period are often of primary interest. Serious losses of goodwill can, however, occur for out-of-warranty infant mortality failures (e.g., a computer that fails one month after a one-year warranty expires).

3.2 Failures due to accidents and other external shocks

Not all failures can be traced to degradation; some product failures are caused by sudden accidents. For example,

- A tire may be punctured by a nail in the road.
- A computer modem may fail from a lightning induced charge an unprotected telephone line.

Failures due to external shocks or accidents depend only on the occurrence of the shock and not on the state (i.e., degradation) of the product at the time of the shock. If a failure occurs after multiple shocks, then we would consider it a degradation mode where, for example, the level of degradation would be some measure of accumulated damage.

3.3 Failures due to unavoidable degradation

Some products (or components in systems) exhibit "unavoidable, graceful, degradation toward failure." Examples of degradation failure modes include,

- Tread on automobile tires and friction material on automobile break pads and clutches that wear with use.
- Corrosion causes thinning of walls of pipes in a chemical reactor.
- Filament material in a light bulb evaporates over time.
- Fatigue cracks will initiate and grow in a steel frame if there are sufficiently high bending stresses.

Key concerns here are factors affecting degradation rates and the relationship of degradation to actual failure events. It may be possible to extend life of degrading components by reducing the rate degradation, by designing the product to be more robust to degradation, or by replacement of degrading components before they can fail in service.

3.4 Comments

We can see that the boundaries among these types of failures are not always sharp. For example, an accident could set up the conditions needed for degradation to progress or a product may become more susceptible to accidental failure after some degradation. Nevertheless, the structure will be useful in our subsequent discussion.

"Surprise" or unanticipated failure modes generally cause the most serious (i.e., expensive) reliability problems. Some failure modes are discovered only after a product has been introduced into actual service. Many unexpected failure modes are traced to a defect-causing degradation mechanism (e.g., unexpected corrosive elements in an operating environment that were not present during laboratory life tests). Others are due to accidents, usually, where sufficient protection was not provided (e.g., damage to electronics from lightning or electro-static discharge).

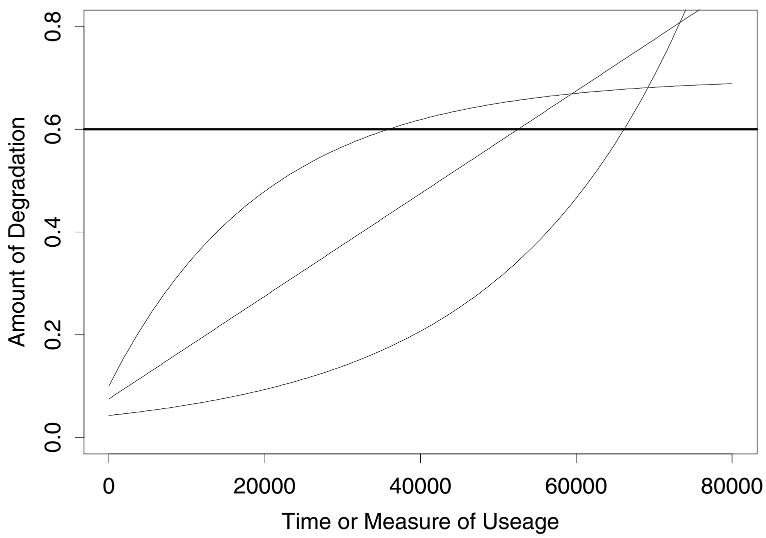

Figure 2: Examples of possible shapes for univariate degradation curves using arbitrary units of degradation and time

4 A Generic Model for Degradation-Caused Failures

4.1 Degradation leading to failure

Most failures can be traced to an underlying degradation process. Figure 2 shows examples of three possible shapes for univariate degradation curves in arbitrary units of degradation and time: linear, convex, and concave. The horizontal line at degradation level 0.6 represents the level or approximate level at which failure would occur. In many applications there may be more than one degradation measure or more than one underlying degradation process. For the discussion in this paper, however, we will restrict discussion to failure modes with a single degradation process and corresponding degradation measure.

Linear degradation might be expected in wear processes where the rate of wear is constant over time (e.g., tire wear). Linear degradation models were used in [24] to model the increase in a resistance measurement over time and in [27], modeling lumens output from fluorescent light bulbs over time.

Degradation curve models in which the rate of growth tends to increase with the level of degradation are used in modeling the growth of fatigue cracks. For example, [16] use the Paris model for the growth rate of fatigue cracks. Letting a denote the size of a crack, a measure of degradation, the Paris model (e.g., [5]), says that

$$\frac{da}{dt} = C\left[\Delta K(a)\right]^m, \tag{1}$$

where C and m are materials properties and $K(a)$ is a function of crack size a which also depends on applied stress, part dimensions, and geometry. The convex curve in Figure 2 is the solution to this differential equation with fixed C, m, and $K(a)$ function.

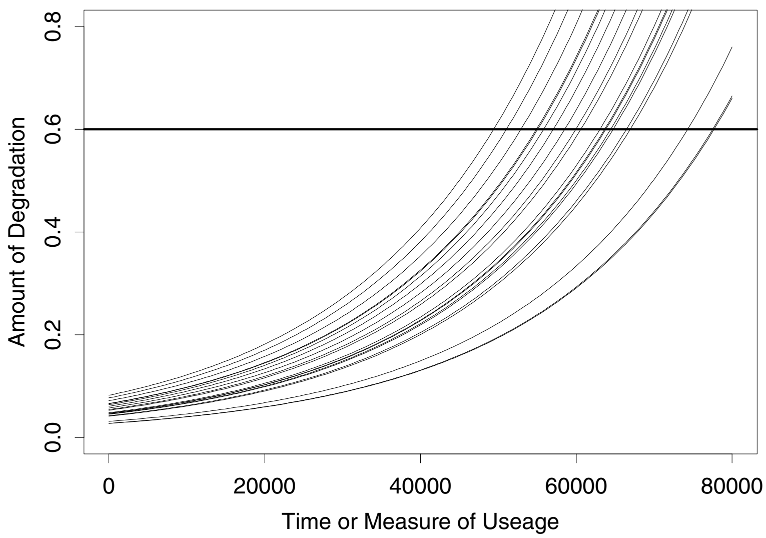

Figure 3: Plot of Paris model for growth of fatigue cracks with unit-to-unit variability in the initial crack size but with constant materials parameters (C and m), and constant stress

Models in which the rate of growth decreases with the level of degradation are used in [3] and [21] to describe degradation of components in electronic circuits. For example, the concave curve in Figure 2 depicts Model 1 from [21]. This curve is a graph of

$$A_2(t) = A_2(\infty) - [A_2(\infty) - A_2(0)]\exp(-k_1 t), \tag{2}$$

the solution to a system of differential equations based on chemical kinetics to predict $A_2(t)$, the amount of failure-causing chlorine compounds as a function of time. The asymptote at $A_2(\infty)$ is a function of the limited initial amount of reactive compounds containing chlorine and available for reaction to the harmful compounds.

4.2 Models for variation in degradation and failure time

If all manufactured units were identical, operated at exactly the same time, under exactly the same conditions, and in exactly the same environment, and if every unit failed as it reached a particular "critical" level of degradation, then all units would fail at exactly the same time. Of course, there is some degree of variability in all of these failure-related factors and these combine to cause variability in failure times.

4.2.1 Unit-to-unit variability

The following sources of variability are due to variability in the characteristics of individual units.

- **Initial level of degradation or amount of harmful, degradation-causing material.** Figure 3 shows the Paris model for growth of fatigue cracks, with simulated variability in the size of the initial crack, but with all other of the unit's characteristics and other factors affecting crack growth rate held constant.

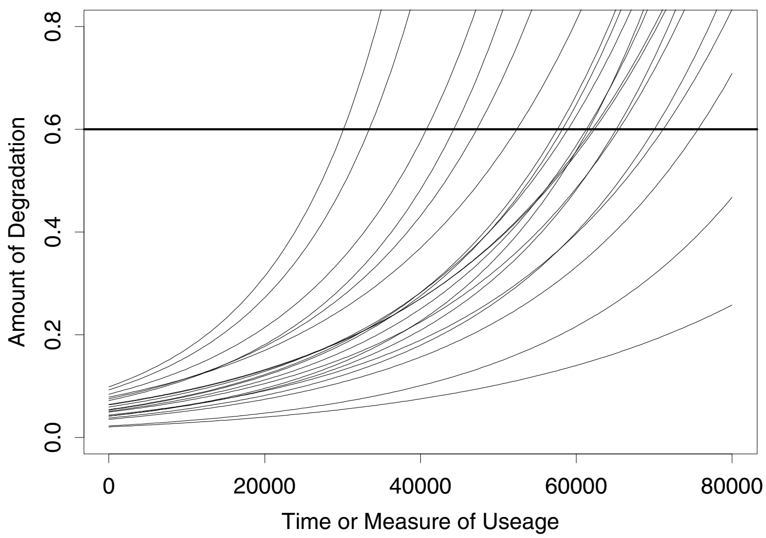

Figure 4: Plot of Paris model for growth of fatigue cracks with unit-to-unit variability in the initial crack size and materials parameters (C and m), but with constant stress

- **Material properties**. Figure 4 shows the Paris model for growth of fatigue cracks, allowing for unit-to-unit variability in the material properties parameters C and m and the size of the initial crack. In this case, as shown in the Paris model in (1), the rate of growth depends on C and m and this additional variability allows for crossing of the crack-growth curves (typical of what is observed in actual fatigue testing). Figure 5 shows simulated curves from the model in (2) depicting the amount of A_2, the failure-causing chlorine compounds, as a function of time. The curves illustrate variability in $A_2(0)$, k_1, the reaction rate, and in the asymptote, $A_2(\infty)$.

- **Component geometry or dimensions.** Unit-to-unit variability in component geometry or dimensions can, for example, cause additional unit-to-unit variability in degradation rates [as in the Paris model in (1)].

Taguchi [25] calls these sources of variability, due mostly to manufacturing or raw materials variability, "unit-to-unit noises."

4.2.2 Variability due to operating and environmental conditions

Besides the materials properties described in Section 4.2.1, the rate of degradation toward failure will depend on operating and environmental conditions. Taguchi [25] calls these sources of variability "external noises." For example, the $K(a)$ in the Paris model (1) depends on the amount of applied stress. In laboratory fatigue tests, the amount of stress is either fixed or changing in a systematic manner (e.g., to keep $K(a)$ nearly constant as a increases). In actual operation of most components, stress would generally be a complicated function of time, described by a stochastic process model. Figure 6 shows the Paris model example with degradation rate at each instant (actually over intervals of size 400 miles)

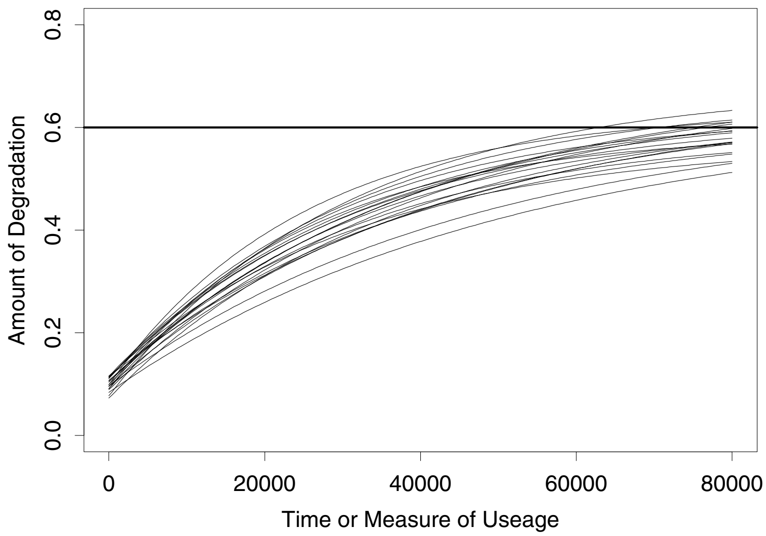

Figure 5: Plot of limited-level degradation model in which the initial conditions, growth rate, and limiting level vary from unit to unit

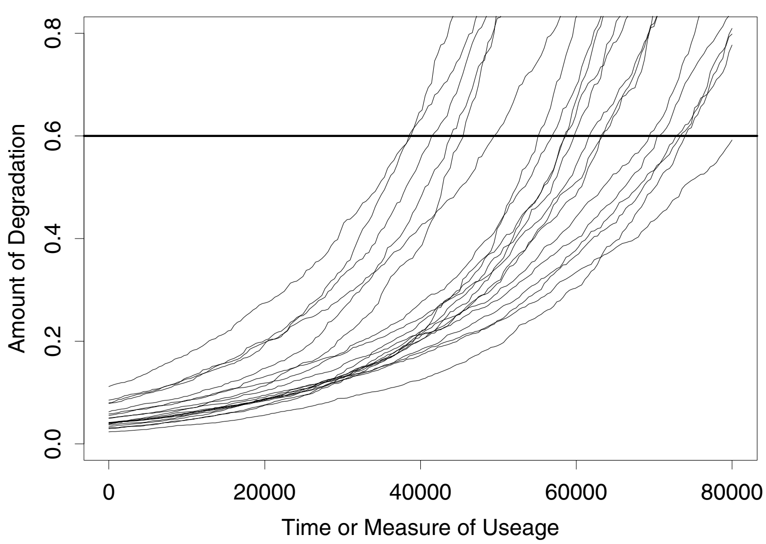

Figure 6: Plot of Paris model for growth of fatigue cracks with unit-to-unit variability in the initial crack size and materials parameters (C and m), and with a stochastic process model describing the changes in stress over the life of the unit

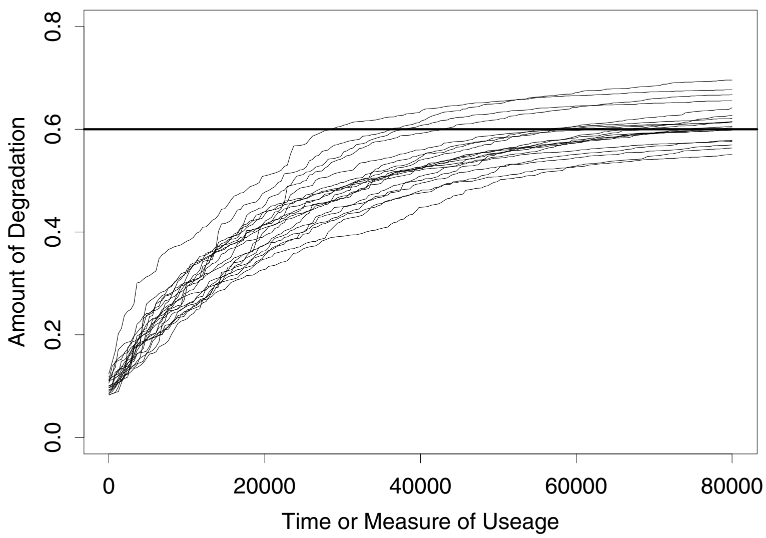

Figure 7: Plot of limited-level degradation model in which the initial conditions, growth rate, and limiting level vary from unit to unit and with a stochastic process describing a time-varying degradation rate within each unit

determined by a white-noise stochastic process. For many chemical degradation processes, the degradation rate will depend on factors like temperature and humidity. Figure 7 shows the failure-causing chlorine compound model in (2) with variability simulated by a stochastic process (sine-wave plus white noise) for the degradation rate k_1. The resulting rough curves are typical of what is seen in actual data with variations in operating temperature causing corresponding variation in degradation rates.

Environmental factors play an important part in the determination of product reliability. Automobiles corrode more rapidly in geographic areas that make heavy use of salt on icy roads. An automobile battery would be expected to last longer in the warm climate of Florida than in the stressfully cold climate of Alaska. Due to increased heat and ultra-violet ray exposure, paints and other coating materials degrade more rapidly in southern part of the United States. Driving automobiles on poorly maintained roads will cause fatigue failures of certain components to occur more rapidly than would be expected on smooth roads. Electronic components installed in the engine compartment of an automobile are subjected to much higher failure-causing heat, humidity, and vibration than are similar components installed in an air-conditioned office. Closely related is the effect of harsher-than-usual handling or operation of a product. For example, some household sump pumps have only a 50% duty cycle. If such a pump, in an emergency situation, has to run continuously, the temperature of the electric motor's components will become exceedingly high and the motor's life will be much shorter than expected. Excessive acceleration and breaking of an automobile will lead to excessive wear on break pads, relative to number of miles driven.

A large proportion of unexpected reliability problems result from unanticipated failure modes caused by environmental effects that were not part of the initial reliability-evaluation program. When making an assessment of reliability it is important to consider environmental effects. Data from designed experiments or field-tracking studies can be used to assess the effect that environmental and operational variables will have on reliability.

In some applications it is possible to protect products from harsh environments. Alternatively, products can be designed to be robust enough to withstand the harshest expected environments. Generally this can be done by,

- Designing to keep degradation rates low (keeping the degradation curves in Figure 7 from reaching the horizontal line at the critical degradation level of 0.6).
- Designing to allow product to operate with higher levels of degradation (e.g., raising the critical level of degradation level above 0.6 in Figure 7).

Such practices might increase product cost, but could be expected to have exceedingly high reliability in more benign environments. One of the challenges of product design is discover and develop economical methods of building in robustness to environmental and other factors that manufacturers and users are unable to control.

4.3 Relationship between degradation and failure

4.3.1 Hard failures: correlation between failure and degradation level

For some products, the definition of the failure event is clear—the product stops working (e.g., when the resistance of a resistor deviates too much from its nominal value, causing the oscillator in an electronic circuit to stop oscillating or when an incandescent light bulb burns out). We call these "hard failures."

With hard failures, failure times will not, in general, correspond exactly with a particular level of degradation (like the horizontal line shown in Figures 2 through 6). Instead, the level of degradation at which failure (i.e., loss of functionality) occurs will be random from unit to unit and even over time. For example, degradation might represent declining strength and failure occurs when randomly applied stress exceeds strength, as shown in Figure 8. It is possible that product-design factors could affect the rate of degradation as well as the level of degradation at which a hard failure would occur.

4.3.2 Soft failures: specified degradation level

In some product there is a gradual loss of performance (e.g., decreasing light output from a fluorescent light bulb). Then failure would be defined (in a somewhat arbitrary manner) at a specified level of degradation (e.g., 60% of initial output). We call this a "soft failure" definition.

5 Degradation Data and Failure-Time Data

Traditionally, most statistical studies of product reliability have been based on time-to-failure data. For some reliability tests, however, it is possible to record the actual level of degradation on units as a function of time. Degradation data provide sample paths of degradation level as a function of time, as illustrated in Figures 3 to 7. In this section we describe some of the important differences between these different kinds of data.

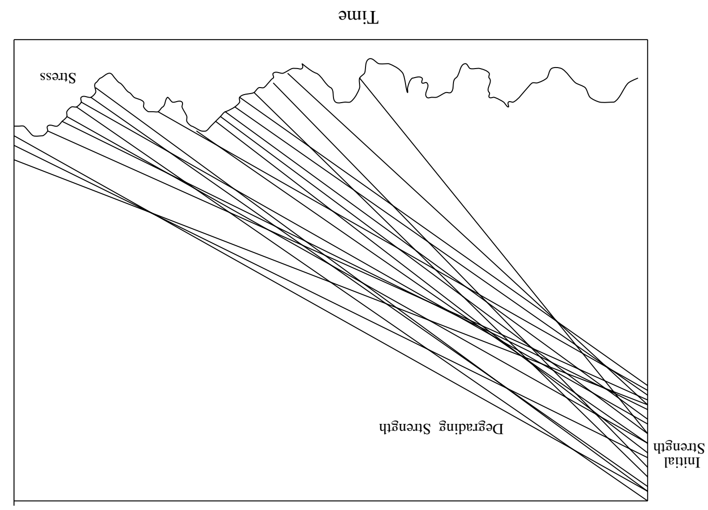

Figure 8: Dynamic stress-strength where strength decreases until random applied stress exceeds strength

5.1 Degradation data

In some reliability studies, it is possible to measure degradation directly over time, either continuously or at specific points in time. In most reliability testing applications, degradation data, if available, will have important practical advantages. In particular,

- Degradation data can, particularly in applications where few or no failures are expected, provide considerably more reliability information than would be available from traditional time-to-failure data.

- Accelerated tests are commonly used to obtain reliability test information more quickly. Direct observation of the degradation process (as opposed to observing a level of performance output) allows direct modeling of the failure-causing mechanism, providing more credible and precise reliability estimates and a firmer basis for extrapolation.

Degradation or performance level is the natural output of many testing processes (e.g., monitoring electrical characteristics of an electronic device). Often, however, degradation measurement requires destructive inspection (e.g., destructive strength tests) or disruptive measurement (e.g., disassembly of a motor, having the potential to change the degradation process). The advantages of degradation data can also be compromised when the degradation measurements are contaminated with a substantial amount of measurement error or if the degradation measure is not closely related to failure.

Degradation curves like those shown in Figure 9 are sometimes seen, especially in monitoring of electrical parameters of electronic devices. A possible reason for the rapid change in degradation is an *unobserved* sudden change in the physical state of the unit that would

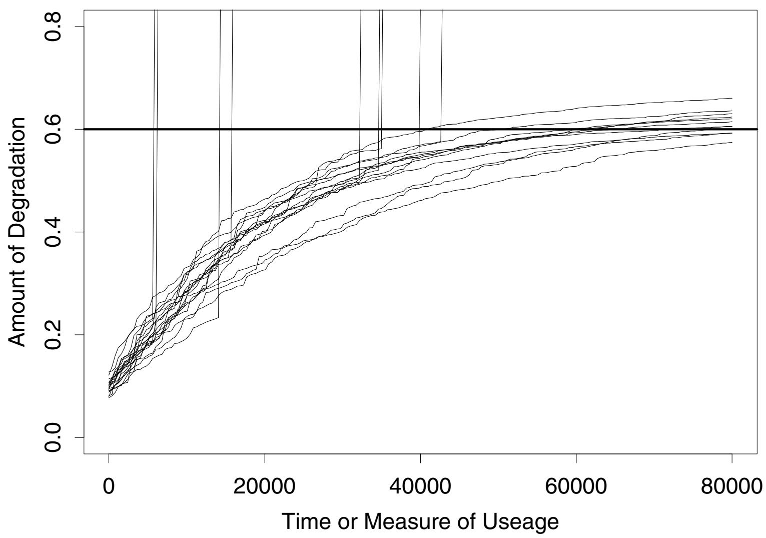

Figure 9: Plot of limited-level degradation model in which a proportion of units fail for a reason that may not be closely related to the measured "degradation" variable.

lead to a subsequent increase in the degradation rate. In such cases the correlation between failure times and early-life degradation will be small and there may be little to be gained by using such a degradation measure instead of traditional failure time data. The limited amount of information in such degradation measurements can be the result of monitoring a *performance* variable (e.g., output voltage) rather than the actual physical degradation (e.g., amount of material displaced by electromigration). An important but difficult challenge of degradation analysis is to find and develop methods for accurately measuring variables that are closely related to failure time.

5.2 Failure time data

For many reliability tests it is difficult, expensive, or impossible to obtain degradation measurements and only (censored) time-to-failure data will be available. Such data can still be used for purposes of reliability assessment and reliability improvement, but with important limitations in the amount of information that will be available:

- As shown in [17], variances of important reliability characteristics like estimates of percentiles of the life distribution will be larger than they would be with degradation data.
- Verification of acceleration model assumptions is much more difficult.
- It will be difficult or impossible to estimate important reliability characteristics like degradation rate.

6 Upstream Accelerated Tests for Materials and Components

Accelerated tests subject test units to higher than usual stresses to obtain life test or degradation information more quickly. Accelerated life tests (observed time to failure) and accelerated degradation tests (observed degradation as a function of time) are used to obtain timely information about reliability/durability/life characteristics of materials or relatively simple system components. In order to obtain information more quickly, units or test specimens are put on test at higher than usual levels of stress (temperature, voltage, pressure, vibration, ultra violet ray exposure, etc.). Then through the use of a statistical model (usually based on kinetics of the underlying degradation process) it is possible to make extrapolative inferences about reliability/durability/life characteristics at lower use or design-levels of the stress variables. Generally, accelerated tests focus on one or a small number of degradation processes or well-understood failure modes.

As indicated in Section 2.1, inferences about life at use or design stress conditions require extrapolation in stress. Such extrapolation is dangerous, but unavoidable in accelerated testing. The magnitude of the inevitable extrapolation error (due to the use of an approximate model) depends strongly on the adequacy of the assumed model and on the *degree* of extrapolation. Meeker and Hahn [20] emphasize this point and suggest how to design accelerated tests to balance the degree of extrapolation with statistical precision.

The justification for extrapolation has to come from the use of reliable models that have been developed from a combination of physical and chemical theory and extensive previous experience in testing similar products and materials. In particular some accelerated test models assume that

- A physical/chemical process causes change in the test unit and that this change progresses over time to eventually cause failure.
- The applied stress or stresses accelerate reaction rates and that this acceleration can be described by a model that is adequate over the range of testing and interest.

The well-known and commonly used acceleration models (e.g., the Arrhenius model for temperature acceleration) have gained acceptance for some applications because of their many successful applications and general agreement of laboratory test results with long-term field performance.

Nelson [22] provides a detailed treatment of the practical and statistical aspects of accelerated tests. Meeker and Escobar [19] review recent research and describe various issues in accelerated testing.

7 Upstream Robust-Design Experiments

Robust-Design is an experimental strategy to make a quality characteristic insensitive or robust to various noise factors. Here “design” refers to the design of the product (although there will also be “experimental designs” to specify how an experiment will be conducted to obtain needed information). For example, using a power circuit output voltage (with a target value) as the quality characteristic of interest, Taguchi, on page 74 of [25], states

"The ideal in functional quality is for the functions to remain normal despite fluctuations in temperature, humidity, supply voltage, and other external (environmental) factors, even when components and materials degrade or wear down during long use, and notwithstanding unit-to-unit variability." Taguchi classifies noise factors into the following categories: (a) manufacturing variation, (b) various environmental and use conditions, and (c) degradation or wear in components and materials (see Section 4.2).

Multifactor robust-design experiments (RDE) provide methods for systematic and efficient reliability improvement. These experiments, conducted on prototype units and subsystems and focus on failure modes involving interfaces and interactions among components and subsystems. Among many possible product-design factors that may impact a system's reliability, RDE's empirically identify the important ones and find levels of product-design factors that yield quality and reliability improvements. For further information see, for example, [4], [10] and [23].

7.1 Effect of product-design factors on reliability

Because degradation occurs over time, one goal of robust-design is to achieve quality over time (i.e., reliability). Product-design factors affect the elements of the degradation-caused failure model of Section 4. In particular, to improve reliability we would want to choose combinations of product-design factors to do one or more of the following:

(a) Decrease the standard deviation and the mean of the degradation rate distribution.

(b) Decrease the standard deviation and increase mean of the critical failure limit distribution. This does not apply if the critical failure limit is user defined, e.g., for fluorescent lamps, when light intensity falls below 60% of light intensity at 100 hours.

(c) Decrease the standard deviation and the mean of the degradation asymptote distribution. This is applicable for limited-degradation models like the one illustrated in Figure 7.

(d) Decrease the standard deviation and the mean of the initial degradation level distribution.

The degradation rate, critical failure limit, degradation asymptote, and initial degradation level described in (a)-(d) are random so that they vary from unit to unit. Furthermore, the quantities in (a)-(c) may change dynamically over the life of a product. To achieve robustness of the system quality characteristic to variations in the noise factors over time, we suggest an extension of the traditional RDE's that will focus on (a)-(d) and, when needed, consider changes over time.

7.2 Static robust-design experiments

Static experimentation implies that individual experimental trials are conducted over relatively short periods of time so that time-related physical change is not observed. Static RDE's can be used to investigate, in addition to other noise factors, the effect of component degradation on system functionality, quality and reliability. This can be done by using, experimentally, components with different values of a critical parameter that might change with time (e.g., resistance, strength, or chemical composition) simulating degradation (or change) that would be expected over time. The response in the experiment is the quality characteristic or characteristics of interest. In terms of the degradation-caused failure model of Section 4, the primary purpose of static RDE's is in increasing "robustness" by reducing the effect that degradation will have, over time, on a system's quality characteristic or eventual failure (e.g., by raising the level of degradation that can be tolerated).

An important advantage of static RDE's is that they can be done quickly, without having to wait for actual degradation or failures to occur. Not having to wait for degradation or failures allows the use of a confirmatory experiment to verify directly that the predicted optimum product-design factor-level combination will, in fact, provide reliability improvement. For some examples of static RDE's see [2] and [23].

Static RDE's, however, have limitations.

- It may be difficult or impossible to obtain units with simulated degradation needed to assess robustness to levels of degradation on the quality characteristic of interest.
- To predict reliability, degradation rates and patterns must be known ahead of time (e.g., from previous experience or accelerated tests).
- In some products the primary focus is in reducing the degradation rate (e.g., the degradation of light intensity of fluorescent lamps or the wear of tires) and this can only be assessed over time.
- Static RDE's experiments do not capture the joint impact of component degradation and environmental factors. For example, there may be interaction between noise factors and product-design factors that affect the degradation rate.

In such cases, the only recourse is to study the degradation process directly in a dynamic RDE.

7.3 Dynamic robust-design experiments with degradation data

Dynamic robust-design experiments provide information, *over time*, on the effect that levels of product-design factors and noise factors will have on the degradation process or on time-to-failure distribution, or both.

Degradation data can provide rich information on the effect that experimental factors have on the components (a)-(d) of the degradation failure model in Section 7.1. This information can be used to suggest methods for improving reliability. To analyze data from dynamic RDE's with degradation data one can use the observed values of the quantities described in Section 7.1 to estimate the parameters of the random-effects models (e.g., models that allow degradation rates to change from unit to unit), as functions of the experimental factors (e.g., [16]). Given a (possibly probabilistic) relationship between degradation and failure, the degradation model will provide estimates of the failure time distribution characteristics (e.g., estimates of failure probabilities or distribution quantiles).

Section 7.6 of [26] describes an experiment to improve the reliability of passenger car tires in which five product-design factors (tire-core type, grain size distribution of carbon,

carbon quantity, additive type and pressing temperature) are studied. Tires were measured at all four positions on four different cars for tire unbalance just before the tires were put on the car and after one and two months. The purpose of the experiment was to determine the levels of the five product-design factors that will result in minimal degradation in tire unbalance across different levels of tire position and car; that is, tire position and car can be regarded as noise factors.

In [27] there is a study of the effect of three product-design factors on the degradation of fluorescent lamp intensity over a period of time. This experiment, however, did not include noise factors. Because failure was defined in terms of percent degradation from a unit's initial condition, only the degradation rate distribution was of interest. Two of the three product-design factors were found to affect the degradation rate and the experiment suggested that a 70% increase in the median of the life distribution could be achieved.

7.4 Dynamic robust-design experiments with only failure-time data

As discussed in Section 5, even if there is a degradation process leading to failure, it may be difficult or impossible to measure the amount of degradation. In such situations, *failure time data* can still provide useful information for reliability improvement. In [11] and [12] Hamada reviews a number of dynamic RDE's both with and without noise factors in which failure-time was the reported response. It is common to see experiments with failure time as the reported response, when a suitable, more informative degradation measure could easily have been recorded as well.

7.5 Accelerated robust-design experiments

Generally degradation proceeds slowly under use conditions making it difficult to get timely information from a dynamic RDE. The problem is especially acute when only failure time data are available. In such cases, it might be possible to use ideas from accelerated testing to obtain information more quickly. For example, use-rate might be increased or tests might be run at higher than usual temperature, as described in Section 6. Recall the importance of having a physical acceleration model as discussed earlier because less data is required (i.e., fewer accelerating factor levels) and predictions are more reliable.

Chapter 21 of [4] discusses an experiment to improve the reliability of surface mount capacitors which used acceleration. Other experiments described in [4] employed a regimen of temperature and humidity cycling to accelerate failure.

7.6 Environmental stress/life (STRIFE) prototype tests to identify failure modes

The pressure to quickly develop new, high reliability products has motivated the development of other new-product testing methods. The purpose of these testing methods is to quickly identify and eliminate potential reliability problems early in the design stage of product development. One such testing method is known as STRIFE (STRess-lIFE) testing. The basic idea of STRIFE testing is to aggressively stress and test prototype or early production units to force failures. Although it may be useful to test only one or two units, more replicates will provide important additional information on unit-to-unit variation.

For more information, see [1] which reports an example in which the complete STRIFE test and improvement program was successfully completed in three weeks. Haible [9] describes general ideas behind STRIFE testing and guidelines for conducting STRIFE tests, compares the method with other types of reliability testing, and outlines an example.

Because STRIFE tests result in product design changes, it is, perhaps, pointless to use the resulting data to predict normal use reliability. Nevertheless, ideas from statistical experimental design and models relating stress to life could be useful in choosing stresses, stress ramp speed, and other aspects of the test.

8 Downstream Tests for Reliability Demonstration, Improvement: Burn-in and Environmental Stress Screening

Burn-in tests can be viewed as a type of 100% inspection or screening of the product population to eliminate or reduce the number of defective items that cause infant mortality. All product units are run for a period of time before installation or use. To accelerate the process, components like integrated circuits may be run at high levels of temperature and/or other stresses. The ability to use acceleration is much more limited for systems and subsystems. See [13] for an engineering approach to this subject. Also, see page 43 of [22].

Environmental Stress Screening (ESS) was developed as a means of accelerated burn-in for units at the system or subsystem (e.g., circuit pack) level. ESS uses mild, but complicated stressing such as temperature cycling, physical vibration, and perhaps stressful operational regimes (e.g., running computer CPU chips at higher than usual clock speeds and lower than usual voltages) to help identify the weak units. Tustin [28] provides a motivational description of the methodology and several references. Page 39 of [22] gives additional references, including military standards.

Burn-in and ESS are inspection/screening schemes. In line with the modern quality precept of eliminating reliance on mass inspection, most manufacturers would prefer not to do burn-in or ESS. They are expensive and may not be totally effective. By improving the reliability through continuous improvement of the product design and the manufacturing process, it may be possible to reduce or eliminate reliance on screening tests except, perhaps, in the most critical applications.

9 Field Data

The most costly reliability problems are discovered after product has been introduced into the field. Often, problems are caused by failure modes that were not expected and were not observed in any laboratory testing. There are many possible reasons for "surprise" failure modes, as described in Section 3. Although problems that are first discovered in the field are generally expensive to correct, for the following reasons, it is often important to develop and use a process for systematic acquisition of some field data.

- If a product has what could turn out to be a serious reliability problem, it is important to learn about the problem as early as possible.

- Field information can be used as feedback to improve future generations of the same or similar product. Failure mode and degradation data are particularly rich in information that can be used to suggest methods for reliability improvement.
- Customer impressions and information on environmental and other product-use conditions (particularly those that have the potential to lead to product problems or failure) are useful for reliability improvement programs.

Data from the field can be obtained in many different ways.

- Some manufacturers will establish partnerships with customers who will thoroughly exercise early production units to provide feedback before large numbers of units have been shipped to regular customers. Examples include manufacturers who provide employees with early-production units of their products and software vendors who provide developers with early (beta-test) releases of their software products. This approach is most effective if at least some units are subjected to the most severe operating environments.
- Most products sold under warranty will generate warranty repair data. Generally extra effort is needed to obtain detailed information.
- In some cases, manufacturers will do "field tracking" where they carefully track the performance of a sample of units in the field.

See [14] and [15] for discussion of important issues in the collection and analysis of field reliability data.

10 Concluding Remarks: Defining the Reliability Assurance Process

In today's manufacturing environment, the goal is to design, develop, test, and manufacture high reliability economical products in short product-cycle times. Meeting this goal will require the development of reliability assurance processes that produce, integrate, and use engineering, scientific and empirical reliability information. Figure 1 illustrates the interaction between engineering knowledge and data-information flows in three product-stages: design, manufacturing, and use. Customer requirements, engineering theory, and other knowledge combine to develop robust product and robust process designs. Upstream data and information sources (ALT, ADT, RDE, etc.) for reliability characterization and improvement are tightly linked to this development.

Manufacturing variation can reduce product reliability. Effective process control and process monitoring is important for maintaining high quality/reliability. Similarly, improper assembly, process operating conditions, or handling can lead to undesired consequences. Explicit consideration of design-manufacturability and the development of robust manufacturing processes can enhance product reliability.

In parallel with the modern quality philosophy, manufacturers would rather not rely on expensive downstream reliability screening (burn-in and ESS) to assure product reliability. Upstream efforts can be more effective and less expensive. When downstream reliability testing is needed and used (especially for under sea and space applications where the additional cost of such methods is clearly justified) failure information can be useful in the reliability improvement process.

Careful collection and monitoring of early-production data (both laboratory testing and field use) can provide early warning of potential reliability problems. Although reliability problems found at this stage may require expensive remedies, it is *much* better to discover problems earlier rather than later in the production cycle. Field data also provides a rich source of information for future product generations. In either case, potentially important auxiliary information (use rates, environmental conditions, amount of degradation, etc.) on both surviving *and* failing units should be collected and fed-back to design engineers.

The design of an effective reliability assurance process may include all or only some of the elements in Figure 1. Particular market and technological conditions will determine the best structure. Meeting today's reliability goals can, however, be expected to require the development of better methods for conducting upstream experiments to provide needed information.

Acknowledgments

We would like to thank Luis A. Escobar, Gerry Hahn, Jerry Lawless, C. Joseph Lu, Jock MacKay, and an anonymous referee for helpful comments on an earlier version of this paper. M. Hamada's research was supported in part by General Motors of Canada Limited, the Manufacturing Research Corporation of Ontario, and the Natural Sciences and Engineering Research Council of Canada.

References

[1] R. A. Bailey, R. A. Gilbert, STRIFE testing for reliability improvement, *Proceedings of the Institute of Environmental Sciences*, 1981, pp 119-121.

[2] D. Byrne, J. Quinlan, Robust function for attaining high reliability at low cost, *1993 Proceedings Annual Reliability and Maintainability Symposium*, 1993, pp 183-191.

[3] M. B. Carey, R. H. Koenig, Reliability assessment based on accelerated degradation: a case study, *IEEE Transactions on Reliability,* R-40, 1991 Dec, pp 499-506.

[4] L. W. Condra, *Reliability Improvement with Design of Experiments*, 1993, New York: Marcel Dekker, Inc.

[5] N. E. Dowling, *Mechanical Behavior of Materials*, 1993, Englewood Cliffs, NJ: Prentice Hall.

[6] R. Evans, Bayes is for the birds, *IEEE Transactions on Reliability,* R-38, 1989 Oct, pp 401.

[7] G. J. Hahn, W. Q. Meeker, Pitfalls and practical considerations in product life analysis, part 1: basic concepts and dangers of extrapolation, *Journal of Quality Technology*, Vol. 14, No. 3, 1982 Oct, pp 144-152.

[8] G. J. Hahn, W. Q. Meeker, Assumptions of statistical inference, *The American Statistician* 47, 1993 Feb, pp 1-11.

[9] C. Haible, Strife testing method of finding relevant reliability problems, Contribution in Technical Report SETEC91-017, 1991, Sandia National Laboratories.

[10] M. Hamada, Reliability improvement via Taguchi's robust design,*Quality and Reliability Engineering International*, 9, 1993 Jan/Feb, pp 7-13.

[11] M. Hamada, Using statistically designed experiments to improve reliability and to achieve robust reliability, *IEEE Transactions on Reliability* R-44, 1995 June, pp xxx-xxx (same issue).

[12] M. Hamada, Analysis of experiments for reliability improvement and robust reliability, in *Recent Advances in Life-Testing and Reliability*, 1995, N. Balakrishnan, editor, Boca Raton: CRC Press.

[13] F. Jensen, N. E. Petersen, *Burn-in, An Engineering Approach to Design and Analysis of Burn-in Procedures*, 1982, New York: John Wiley & Sons, Inc.

[14] J. D. Kalbfleisch, J. F. Lawless, Estimation of Reliability in Field-Performance Studies, *Technometrics*, 30, 1988 November, pp 365-388 (with discussion).

[15] J. F. Lawless, J D. Kalbfleisch, Some issues in the collection and analysis of field reliability data, in *Survival Analysis: State of the Art*, 1992, pp 141-152, J. P. Klein and P K. Goel (eds.), Netherlands: Kluwer Academic Publishers.

[16] C. J. Lu, W. Q. Meeker, Using degradation measures to estimate a time-to-failure distribution, *Technometrics* 34, 1993 May, pp 161-174.

[17] C. J. Lu, W. Q. Meeker, L. A. Escobar, A comparison of degradation and failure-time analysis methods of estimating a time-to-failure distribution, 1995, Department of Statistics, Iowa State University.

[18] W. Q. Meeker, Limited failure population life tests: application to integrated circuit reliability. *Technometrics* **29**, 1987 Feb, pp 51-65.

[19] W. Q. Meeker, L. A. Escobar, A review of recent research and current issues in accelerated testing, *International Statistical Review*, 61, 1993, pp 147-168.

[20] W. Q. Meeker, G. J. Hahn, *How to Plan Accelerated Life Tests: Some Practical Guidelines*, Volume 10 of the ASQC Basic References in Quality Control: Statistical Techniques, 1985. Available from the American Society for Quality Control, 310 W. Wisconsin Ave., Milwaukee, WI 53203.

[21] W. Q. Meeker, M. J. LuValle, An accelerated life test model based on reliability kinetics, *Technometrics* 37, 1995 May, pp xxx-xxx (in press).

[22] W. Nelson, *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, 1990, New York: John Wiley & Sons, Inc.

[23] M. S. Phadke, *Quality Engineering Using Robust Design*, 1989, Englewood Cliffs, NJ: Prentice Hall.

[24] K. Suzuki, K. Maki, S. Yokogawa, An analysis of degradation data of a carbon film and properties of the estimators, in *Statistical Sciences and Data Analysis*, 1993, (K. Matusita, M. Puri, and T. Hayakawa, Editors), Utrecht Netherlands: VSP.

[25] G. Taguchi, *Introduction to Quality Engineering*, 1986, Tokyo: Asian Productivity Organization.

[26] G. Taguchi, *System of Experimental Design,* 1987; White Plains, NY: Unipub/Kraus International Publications.

[27] T. S. Tseng, M. Hamada, C. H. Chiao, Using degradation data from a factorial experiment to improve fluorescent lamp reliability, 1995, to appear in *Journal of Quality Technology.*

[28] W. Tustin, Shake and bake the bugs out, *Quality Progress,* 1990 September, pp 61-64.

AUTHORS

Dr. William Q. Meeker; Department of Statistics, 326 Snedecor Hall, Iowa State University, Ames, Iowa 50011 USA.
Internet (email): wqmeeker@iastate.edu

William Meeker has been a professor in the Department of Statistics at Iowa State University since 1975. He holds a BS from Clarkson College and MS and Ph.D. degrees from Union College. He has worked three summers at GE CRD and 15 summers at AT&T Bell Laboratories. He is a Fellow of the American Statistical Association, winner of the American Society for Quality Control Wilcoxon Prize, and an elected member of the International Statistics Institute. His interests are in the areas of reliability, reliability data analysis, accelerated testing, and statistical computing. He is author of two books, and of numerous publications in the engineering and statistical literature; he is a former editor of *Technometrics*.

Dr. Michael S. Hamada; Department of Statistics and Actuarial Science and The Institute for Improvement in Quality and Productivity; University of Waterloo, Waterloo, Ontario N2L 3G1 Canada.

Internet (email): mshamada@watstat.waterloo.edu

Michael S. Hamada is Research Associate Professor in the Department of Statistics and Actuarial Science and The Institute for Improvement in Quality and Productivity at the University of Waterloo. His research interests include design of experiments, reliability and statistical process control. He received a Ph.D. in Statistics from the University of Wisconsin–Madison in 1987.