7-1-2003

Reliability: The Other Dimension of Quality

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

Luis A. Escobar
Louisiana State University

Reliability: The Other Dimension of Quality

Abstract
During the past twenty years, manufacturing industries, particularly in the United States, have gone through a revolution in the use of statistical methods for product quality. Tools for process monitoring and, particularly experimental design, are much more commonly used today to maintain and improve product quality. A natural extension of the revolution in product quality is to turn focus to product reliability, which is defined as "quality over time." This has given rise to programs like Design for Six Sigma.

In this paper we discuss the relationship between engineering quality and reliability and outline the role of statistics and statisticians in the field of reliability. We provide a brief introduction to the statistical tools used in engineering reliability and make some predictions for the future of statistics in engineering reliability.

Keywords
Demonstration Test, Maximum Likelihood, Reliability Assurance, Reliability Data, Robust Design, Warranty Data

Disciplines
Statistics and Probability

Comments
This article was published in *Quality Technology & Qualitative Management* 1 (2004): 1–25.

Reliability: The Other Dimension of Quality

William Q. Meeker
Department of Statistics
Iowa State University
Ames, IA 50011
wqmeeker@iastate.edu

Luis A. Escobar
Department of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803
luis@lsu.edu

July 1, 2003

Abstract

During the past twenty years, manufacturing industries, particularly in the United States, have gone through a revolution in the use of statistical methods for product quality. Tools for process monitoring and, particularly experimental design, are much more commonly used today to maintain and improve product quality. A natural extension of the revolution in product quality is to turn focus to product reliability, which is defined as "quality over time." This has given rise to programs like Design for Six Sigma.

In this paper we discuss the relationship between engineering quality and reliability and outline the role of statistics and statisticians in the field of reliability. We provide a brief introduction to the statistical tools used in engineering reliability and make some predictions for the future of statistics in engineering reliability.

Key words: Demonstration Test, Maximum Likelihood, Reliability Assurance, Reliability Data, Robust Design, Warranty Data.

1 Introduction

1.1 Background

Today's manufacturers face intense global competition, pressure for shorter product-cycle times, stringent cost constraints, and higher customer expectations for quality and reliability. This combination raises some formidable engineering, managerial, and statistical challenges.

1.2 Reliability

The usual textbook definition of reliability reads something like "the probability that a unit will perform its intended function until a given point in time under *specified* use conditions." A more appropriate definition is "the probability that a unit will perform its intended function until a specified point in time under *encountered* use conditions." The point is that the environment in which a product operates is a critical factor in evaluating a product's reliability.

Condra (1993) states "Reliability is quality over time." This implies that good quality is necessary but not sufficient! One major difficulty and major contrast between quality and reliability is that reliability can be assessed directly only after a product has been in the field for some time; accurate reliability prediction presents a number of technical challenges.

Reliability is an engineering discipline. Statistical methods are, however, important tools for reliability engineering. Historically, most statistical effort has been on the development of methods for assessing reliability. Much engineering effort is (correctly) focused on reliability improvement. Only recently have statisticians begun to have an impact on *improving* reliability.

1.3 Engineering functions that affect reliability

In product design, engineers have the following responsibilities (among others):

- Define product requirements.
- Design the product.
- Verify product design.
- Improve the design to assure product robustness.

Then there is a parallel set of steps for manufacturing *process* design. The ideas presented here will generally apply to both engineering product design and engineering process design. After manufacturing begins, there are generally on-going efforts to:

- Improve quality and reliability through design changes.
- Reduce costs through design changes.
- Maintain quality in production (e.g., through process monitoring).

Collectively, these efforts might be called "continuous improvement."

1.4 Overview

The remainder of this paper is organized as follows. Section 2 describes some of the basic ideas behind the common practices of establishing reliability targets and making reliability predictions. Section 3 contrasts traditional reliability demonstration with the more modern concept of reliability assurance. Section 4 explains the relationship between quality and reliability and the negative role that variability plays in each. Section 5 describes methods for making a product less sensitive to environmental noises (variability). In Section 6 we briefly describe some of the different kinds of reliability data and models that are used both for laboratory life tests and for field data. Section 7 introduces some of the important ideas relating to warranty data. Section 8 comments on current practice with respect to statistics and statisticians in the area of product reliability. Concluding remarks are given in Section 9.

2 System reliability targets, models, and prediction

2.1 Motivation

Design for Reliability requires careful consideration of product (process) failure modes. Broadly, failure modes can be classified as those that are anticipated and those that are unanticipated.

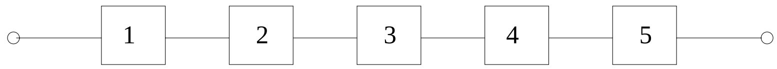

Figure 1: System reliability diagram for a system with five components in series.

Generally a reliability model will reflect only the anticipated failure modes and it is the anticipated failure modes that engineers focus on (although it can be argued that design for product robustness will help to prevent even unanticipated failure modes). The focus of our discussion will be on product design and product reliability.

2.2 System reliability targets and reliability prediction

Manufacturers typically have formal or informal reliability goals for their products. Such goals are generally derived from past experience with similar products, industry standards, customer requirements, or a desire to improve an existing reliability of a product. For example, one automobile manufacturer has stated that no more than 10% of their products should experience a serious failure in the first ten years of life. This would imply a reliability target of 0.9.

There has been a considerable amount of concern expressed with the common practice of setting reliability targets for products. The primary concern is that there may be complacency if it is felt that a target has been met, forgoing the opportunity for reliability improvement. Most reliability practitioners would agree, however, that reliability targets, themselves, are not the problem. The problem is in the manner in which they might be misused.

In many product design applications, engineers try to use a predictive model that will provide at least a rough idea of a product's reliability in the field and to assess if reliability targets are being met. Although widely practiced, the development and use of reliability models is also controversial. To some degree this is because of numerous examples where such models were developed and trusted for making important decisions, but then were, in the end, proved to be seriously inaccurate. A more constructive view of these model failures is that they are important parts of the learning process for an important task.

Davis (1998) argues that reliability prediction is extremely difficult and inadvisable because there is, for most products, only a poor characterization of the products use environment. Again he (and also see Grove and Davis (1992) recommends that efforts should focus on reliability improvement, rather than spending scarce resources to simply assess and predict reliability.

2.3 Anticipated failure modes, system reliability models, and reliability budgeting

After engineers have identified all anticipated failures modes, often by using a Failure Modes and Effects Analysis (FMEA) process, they can develop a system reliability model for the *anticipated* failure modes. For example, Figure 1 is a reliability model diagram for a system with five components in series. In a series system, all components of the system must be operating for the system to operate. Numerous books have been written on the subject of system reliability, giving details on different system configurations and on models and methods for computing system reliability as a function of component reliability. The book by Høyland and Rausand (1994), for example, discusses these and related topics.

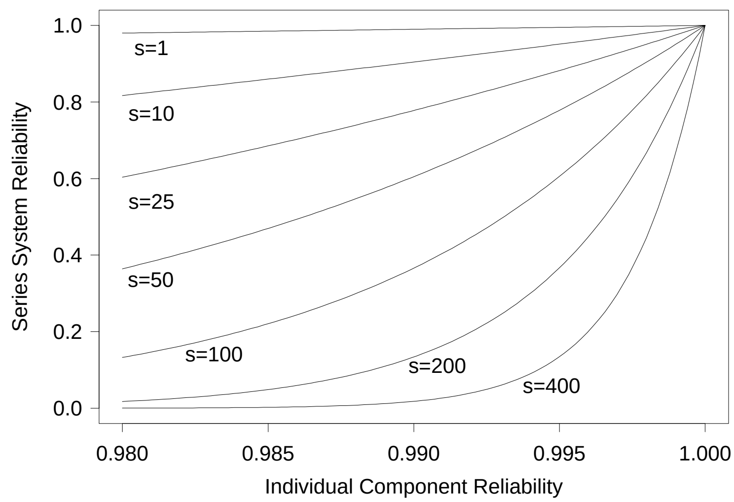

Figure 2: Reliability of a System with s Identical Independent Components in Series.

For a simple series system with s independent components, like the one depicted in Figure 1, the failure time distribution of the system is simply the distribution of the minimum failure time of the s components.

$$F(t) = 1 - \prod_{i=1}^{s} [1 - F_i(t)]$$

where $F_i(t)$ is the failure time distribution of component i. Of course, reliability $R(t) = 1 - F(t)$.

One aspect of the series system model is that in order to have high system reliability, individual component reliabilities have to have extremely high reliability. For example, for the five component system, if all components have reliability 0.9, the system reliability will be only $0.9^5 = 0.59$! Figure 2 (from Chapter 15 of Meeker and Escobar 1998) shows this effect more generally. These curves illustrate the difficulty of achieving high reliability as product part count increases. Indeed, one of the reliability engineer's rules of thumb is to "keep the part count down."

One part of product design is typically to draw up a "reliability budget" that specifies a reliability target for each component such that the system will meets its reliability target. Generally achieving high reliability is much easier for some components than it is for others.

2.4 Unanticipated failure modes

A system reliability model accounts only for the failure modes that engineers anticipate and put into their reliability model. Often, however, it is the unanticipated failure modes that cause the most serious and costly reliability problems. It is for this reason that an important part of any Reliability by Design program should have as one of its goals to discover and eliminate potentially important failure modes as early as possible. Tools for identifying failure modes include (roughly in order of increasing cost) engineering knowledge and previous experience, FMEA in up-front design, highly accelerated life testing (known as "HALT" tests), and early feedback from the field. HALT tests subject prototype subassembly units or systems to higher than usual operating/environmental conditions in order to shake-out design weaknesses. McLean (2000) describes the use of HALT testing, in detail. HALT tests (variations of which go by other names, life Stress-Life testing) are primarily engineering tests to discover failure modes and, as such, provide little data that can be used to quantify or predict product reliability (generally because the discovery of a new failure mode results in a design change that would render past data uninformative for the new design). Such tests are, however, an important part of many engineering product design processes.

Some products undergo "beta testing," where early-production units of a product are released for use, preferably in a high-use, friendly-customer environment (e.g., testing washing machines in a laundromat).

Generally, the longer that it takes to identify a failure mode, the more expensive it is to fix it. This implies that it is poor practice to rely on data from the field to discover potential failure modes and that considerable effort in this direction is needed early in the design process.

2.5 Inputs to reliability models

As described above, the inputs for a reliability model include:

- An identification of failure modes.
- A system structure model in which there is a "component" corresponding to each possible failure mode and providing a description on the effect that component failure has on system failure.
- A probability or statistical model providing information about the reliability of the individual components, as a function of the use environment.
- A description of the product use environment (including, for example, distributions of use rates and stresses).

Inputs to a reliability model become more complicated when components have failure time distributions that are not independent. In such cases, the dependency should be modeled (e.g., by using bivariate or multivariate failure time distributions).

In many applications the use environment will be dynamic, changing over time. For example, jet engines experience different amounts of high, medium, and low levels of temperature and physical stress and the effect of such "spectra" of stresses needs to be part of the reliability model.

Determining the needed model inputs under stringent time and economic constraints is always a challenge. Typical sources of information (roughly in order of increasing cost) are:

- Engineering knowledge (e.g., values that can be found in handbooks).
- Previous experience.
- Analysis based physical/chemical models that relate to failure (e.g., chemical kinetic models of degradation or finite element models to describe the effect of stress distributions on failure).
- Physical experimentation and accelerated testing.

When none of the usual sources of information provide the needed information to the desired degree of accuracy, engineers will typically build into the product or process, conservative "factors of safety."

A major challenge in the use of information from a combination of sources is in the development of methods to quantify uncertainty. One promising approach to this problem is the use of "responsible

Bayesian methods." Here we define "responsible Bayesian methods" to mean the use of Bayes methods in which only prior information with a firm basis is included in the analysis. An example of such a procedure is the Los Alamos National Laboratory PREDICT process. For more information, see www.stat.lanl.gov/projects/predict.shtml.

3 Reliability demonstration versus reliability assurance

Managers need to have information about reliability before making product-release decisions. Potential customers need reliability information before deciding to purchase a product. Thus, there is a desire to "demonstrate" product reliability.

3.1 Traditional reliability demonstration

Traditional reliability demonstration is essentially a statistical hypothesis test. It answers the question "do the data provide enough evidence to reject the null hypothesis that reliability is equal to the target." Rejecting the null hypothesis provides a demonstration that the reliability target has been met. Under minimal assumptions one obtains a go no-go assessment of each test unit and uses a binomial distribution to describe the failure probability. Then to demonstrate that reliability at 20,000 cycles is 0.99 (or equivalently that the 0.01 quantile of the failure time distribution $t_{0.01} = 20{,}000$), with 90% confidence, requires testing at least 230 units for 20,000 cycles with zero failures (i.e., $n = \log(\alpha)/\log(1-p) = \log(0.10)/\log(0.99) \approx 230$). To have just an 80% chance of passing the test requires that the true reliability be approximately 0.999 (i.e., $\Pr(\text{pass test}) = 0.999^{230} = 0.794 \approx 0.80$).

The required sample size can be reduced by making certain assumptions (e.g., about the form of the failure-time distribution) and trading a longer test for a smaller sample size. Such assumptions are, however, generally not possible for a system with many failure modes, especially because acceleration of all potential failure modes in an integrated product is difficult or impossible.

3.2 A feasible application of reliability demonstration

Although it is generally impossible to formally demonstrate the desired high levels of reliability for an entire system in a timely manner, it may, under certain circumstances, be feasible to demonstrate reliability for a component or for a system with respect to a particular failure mode. This can be done by making certain assumptions and conducting an accelerated test. As described in Section 10.6.2 of Meeker and Escobar (1998), by assuming that life has a Weibull distribution with a shape parameter of β, a zero-failure test that runs for $k \times 20{,}000$ cycles requires a sample size of

$$n \geq \frac{1}{k^{\beta}} \times \frac{\log(\alpha)}{\log(1-p)}$$

to make the desired demonstration. For the general discussion here, we will assume that $\beta > 1$, suggesting a wearout type of failure mechanism.

For demonstration described in Section 3.1, by assuming that life has a Weibull distribution with a shape parameter of $\beta = 2$, a zero-failure test that runs for $6.77 \times 20{,}000$ cycles will provide the required demonstration with a sample size of only $n = 5$ units. Increasing the sample size n will reduce the length of the test through k and the relationship between n and k depends on β (with larger β, shorter tests result). The probability of passing the test is again the probability of 0 failures which, interestingly, is the same as the Pr(pass test) for the binomial test. That is (again taking the actual reliability to be 0.999),

$$\Pr(\text{pass test}) = 0.999^{\log(\alpha)/\log(1-p)} = 0.999^{\log(0.10)/\log(0.99)} = 0.795$$

Note that Pr(pass test) does not depend on β or n. The tradeoff here is that the Weibull test must be longer and that β must be greater than or equal to the assumed value (e.g., if there are infant mortality failures that could cause β to be smaller than the assumed value) If β is less than the assumed value, the demonstration will be invalid. In some places, it is recommended to take $\beta = 1$, which would be a safe (conservative) assumption if the failure mode is known to be a wearout failure mode.

3.3 Reliability assurance

For complete, complicated, expensive systems, traditional reliability demonstration tests are not practicable. Reliability assurance is the alternative. Reliability assurance is a procedure based on reliability modeling and combining information from various sources. Inputs to a reliability assurance procedure for a system would generally include knowledge of:

- System structure (how components fit together and interrelate).
- Possible failure modes and their effect on system operation.
- Reliability of individual components and interfaces (including software).
- Knowledge of how the product will be used as well as the environment or environments in which the product will be used.
- Benchmark design information from similar, competing products.

3.4 Structured programs for reliability by design

The idea behind "Reliability by Design" is to build quality and reliability into the product at the design phase and to make the product insensitive (or robust) to variabilities in the production and use environments. The ideas are not new. For example, in 1990, a team at ATT Bell Laboratories published a handbook, ATT (1990), describing a suggested procedure for "Reliability by Design." More recently, Design for Six Sigma (or DFSS as implemented at GE) has the DMADV steps: Define, Measure, Analyze, Design, Verify. There are, of course, other similar company-specific reliability improvement programs that mandate the use of up-front analysis to provide a degree of assurance that a company's product will have the needed level of reliability. Simple web searches for "Design for Reliability" and "Reliability by Design" turned up hundreds of hits. Design for Reliability generally implies the use of up-front analysis and testing (as needed) in product and process design to eliminate problems before they occur. This is in contrast to the traditional Build, Test, Fix, Test, Fix, ... approach that is implied by the "reliability growth modeling" approach to reliability management (e.g., MIL-HDBK-189 1981).

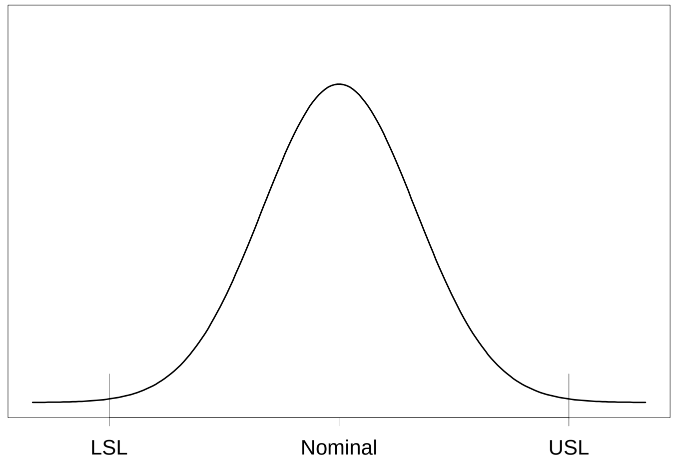

Figure 3: Three-sigma quality characteristic

4 Quality and reliability

4.1 The effect of variability

Figures 3 through 6 illustrate one view of the relationship between quality and reliability. Suppose that the distributions represent some particular product characteristic (e.g., the resistance of a resistor), which will cause degradation of customer-perceived performance (e.g., decreased signal-to-noise ratio) if the characteristic is near to, either inside or outside, of the specification limits. Figure 3 reflects barely acceptable 3-sigma quality. Although the customers whose product is near the center of the distribution may be happy with their product's performance, those closer to the specification limits are less than fully pleased. Over time, as illustrated in Figure 4, there will be drift caused by wear, chemical change, or other degradation, moving more and more customers toward or outside of the specification limits—causing serious reliability problems.

Figure 5 contrasts good quality and poor quality. The point of Figures 4 and 5 is that being within the *specification limits* is not sufficient. Even at time 0, customers near to, but inside of the specification limits may be unhappy with product performance. Assuming that the distribution of the quality characteristic starts in the center of the specification region, smaller variability means that more customers have product close to the target and that these products are more likely to stay within the specification limits over time, providing higher reliability. As illustrated in Figure 6, with good quality, even with expected drift over time, customers will, generally, continue to have good performance—quality over time or high reliability.

4.2 Inspection versus control

Kackar (1989) and Phadke (1989, page 15) describe a costumer preference study published by the Japanese newspaper *Asahi*. The results of that study showed that United States consumers preferred Sony color television sets manufactured in Japan over Sony color television sets manufactured in the

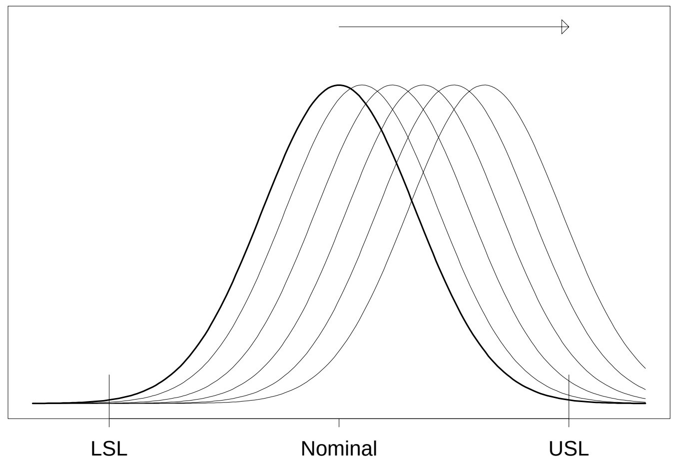

Figure 4: Drifting three-sigma quality characteristic

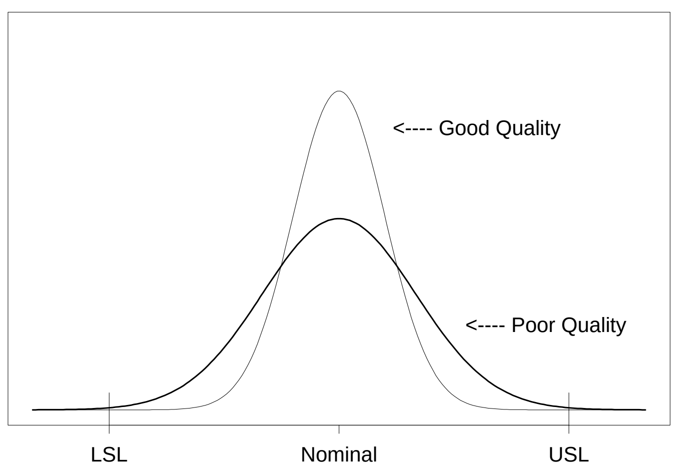

Figure 5: Good and bad quality

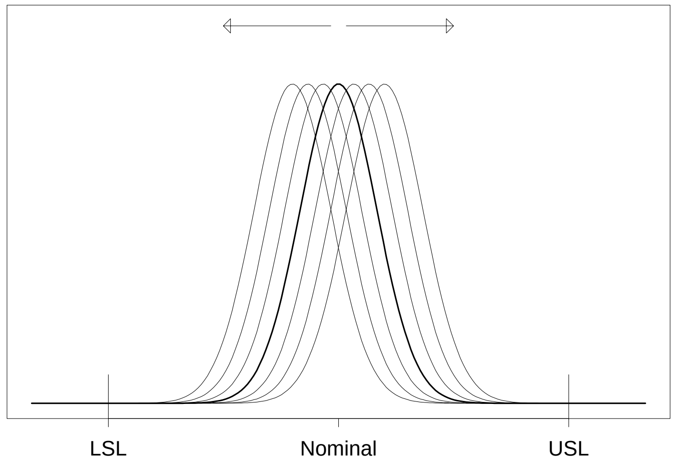

Figure 6: Drifting six-sigma quality characteristic

USA. Initially, this was surprising because both companies had identical design and tolerance limits for manufacturing of the color television sets. Careful analysis showed, however, that Sony Japan's distribution of a particular color-density quality characteristic was highly concentrated about its target with only a small fraction (about 0.3%) of units outside the specification limits. Although Sony USA had almost no units outside the specification limits, the distribution within the specification limits was highly variable. Presumably, Sony Japan was focused on keeping the quality characteristic close to target while Sony USA was focused in meeting the tolerance limits by inspecting out the non-complying products, leading to more scrap, higher manufacturing costs, lower quality, and, ultimately, lower reliability and lower customer satisfaction for Sony USA color television sets.

It is often said that variability is the enemy of quality. Variability is also the enemy of reliability. Examples of important sources of variability are manufacturing (including operators, machine degradation, raw materials), environmental conditions (e.g., stresses, temperature cycling, moisture, dust, and even insects), and customer use rates. Reduction of input variability and reduction in the transmission of input variability to customer perceivable variability are important goals for engineering design.

5 Robust design for improved reliability

5.1 Basic concepts

Robust design is an important, widely known (at least among statisticians working in the area of quality), but still under used concept in quality and reliability. Robustness can be defined as the ability (for a product or a process) to perform its intended function, in an effective manner, under a variety of operating and environmental conditions (including long-term wear or other degradation). Operationally, the challenge is to design a product or process such that it will be robust to the expected environmental "noises" that a product/process will encounter in its manufacture or

Figure 7: Causes of variation in a quality characteristic Y.

operation and to do this in a manner that is economically efficient. While these methods are usually thought of as primarily for the improvement of quality, it is clear that they also have an important role in the improvement of reliability (especially when we recall that reliability is defined as quality over time).

Robust design uses design of experiments during the design stage of a product and its manufacturing process to determine the design parameters that maximize its quality. The operational/technical ideas behind robustness derive from the important engineering ideas that were brought to us by Genichi Taguchi. Taguchi suggested a methodology, based on ideas of statistically designed experiments, that can be used to improve product or process designs by reducing the transmission of variability. The important concepts have been refined and explained by individuals and in places too numerous to mention here, but include, for example, the book-length treatments by Phadke (1989), Grove and Davis (1992), Logothetis and Wynn (1994), Wu and Hamada (2000), and Condra (1993).

Engineering design can be viewed as a complicated optimization problem (although it is not clear how useful it is to do so, given our inability to quantify all of the inputs and especially intangible costs). This begs the question of whether the tools of robust design would be useful for a product or process design that already had been very well engineered (and thus was already very close to optimum). In discussion with statisticians and engineers working in manufacturing, we have been assured, almost uniformly, that few, if any, such designs exist, although clearly this must be product and industry specific.

5.2 Transmission of variability

As suggested by Figure 7, various environmental noises in manufacturing and product use lead to variability in process or product variables (X variables) which in turn cause variability in quality characteristics (Y variables) that are important to the customer.

In terms of probability distributions, this is illustrated with linear transfer functions in Figure 8. Note the interaction between X_1 and X_2 in their relationship with Y. Suppose that X_1 is a "noise" variable that may be difficult or impossible to control in the operation of the product or process (e.g., ambient temperature, use rate, or variable concentration of an input to a chemical process). X_2 is a "design" variable that can be chosen by the product/process designers.

There are, basically, two ways to reduce the variability in Y.

- Reduce the variability in the X variables by, for example, protecting the product from environmental noises or controlling production inputs more carefully (e.g., screening inputs like component characteristics).

- Reduce the transmission of variance through the transfer function.

The former alternative, commonly known as Tolerance Design, is often impossible or unreasonably expensive. Figure 8 suggests, however, that choosing the lower level of X_2 reduces the effect that X_1 variability has on the variability of output Y. By exploiting the interaction between a noise variable

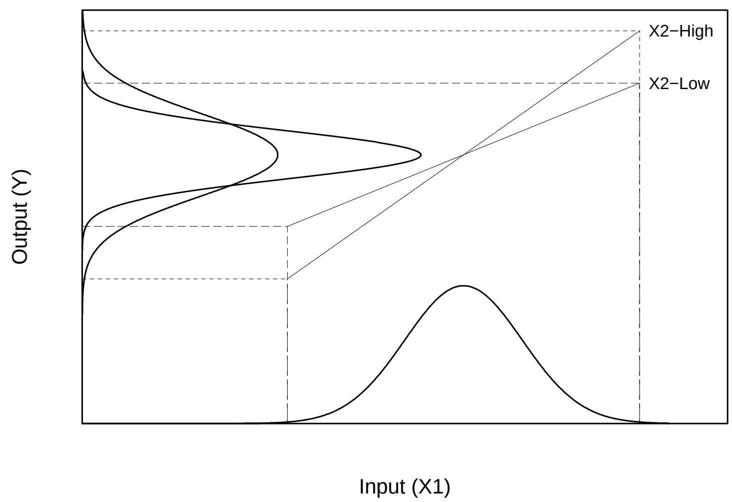

Figure 8: Transmission of variability from an input (noise factor) to an output (quality characteristic)

(X_1) and a design variable (X_2), it may be possible to reduce the variability in Y by making relatively inexpensive changes to the level of X_2.

The experimental strategy suggested by Genichi Taguchi has been refined and explained by others mentioned above. This strategy provides a framework that will allow engineers to identify design variables and settings that will lead to a more robust and reliable product. It should be mentioned that it is not always possible to identify and/or control the noise factors. In such cases designed experiments can still be used in an effort to discover design variables that will have an effect on the variability of the output.

For a more complete discussion of these ideas, see, for example, Chapter 10 of Wu and Hamada (2000).

5.3 Examples

There are numerous examples in the text books mentioned in Section 5.1 and in the engineering literature that illustrate how the methods of robust design can be used to reduce transmission of variability and thereby economically improve reliability. For example, Phadke (1989) describes an electronic circuit design, wherein appropriate choices of the value of a resistor and setting of gain can be used to reduce variability and at the same time keep voltage (a quality characteristic) on target.

Wu and Hamada (2000, chapter 10) describe and present several examples including experiments that were run to improve the reliability of a gear and a leaf spring.

Byrne and Quinlan (1993) show how the reliability of a mechanical speedometer cable was improved by reducing the variability in shrinkage of the cable over time. This was done by appropriate choice of wire braid and wire diameter.

Grove and Davis (1992, Section 8.2) describe the results of an experiment that was used to discover that weld strength (and thus reliability of the part being manufactured) could be increased by simply changing the force with which the pieces of metal were held together while being welded.

In an unpublished internal automobile company technical report, the rate of piston wear was reduced substantially over a mixture of noises variables characterized as operating conditions (notably engine starts in warm and cold environments) by appropriate choices of piston design dimensions and parameters for the lubrication system.

6 Reliability data and models

This section uses a sequence of examples to describe and illustrate some of the important data types and corresponding reliability models and analyzes for the data. Data analyzes were done with the SPLIDA system procedures for life data analysis (see Meeker and Escobar 2003).

6.1 Distinguishing features of reliability data and models

For reliability data analysis, the standard statistical models used in basic statistics courses need to be extended in various directions:

- The normal distribution is rarely used as a model for failure times (instead we use distributions for positive responses such as the lognormal and Weibull distributions).
- Simple moments estimators (means and variances) and ordinary least squares estimators rarely provide appropriate methods of analysis. This is because reliability data are often censored or truncated or have other complications. Instead, simple graphical methods are combined with maximum likelihood estimation for fitting parametric models.
- Model parameters and regression coefficients are not of primary interest. Instead, failure rates, quantiles, probabilities, and reliabilities are needed.
- Extrapolation is often required (e.g., have one year of data, but want proportion failing after three years or have data at high temperatures and need to estimate a failure-time distribution at low temperature).

6.2 Use of physical/chemical models in reliability engineering

As mentioned above, extrapolation is often required in reliability engineering/statistical analyses. Extrapolation is always risky and the basis for extrapolation is generally large amounts of past experience or, preferably, the use of physical/chemical models that describe the physical failure mode mechanisms. Some failure models have been studied extensively over the past decades (for example fatigue in mechanical systems and many mechanisms that cause failures in microelectronics applications). Although the development of such models is challenging, expensive, and time consuming, once the models are available, they will often provide long-term benefits.

6.3 Accelerated testing

Most modern products are designed to operate without failure for years, decades, or longer. Thus few units will fail or degrade appreciably in a test of practical length at normal use conditions. For example, the design and construction of a communications satellite, may allow only 8 months to test components that are expected to be in service for 10 or 15 years. For such applications, Accelerated Tests (ATs) are used widely in manufacturing industries, particularly to obtain timely information on the reliability of simple components and materials. Virtually all laboratory life tests are accelerated in one manner or another.

Consider the following methods of accelerating a reliability test:

- Increase the use-rate of the product. For example, in a life test of a toaster component, one might test toasters 200 times each day. Assuming a typical use profile of 2 uses per day, a test running 12 days could estimate the life distribution out to 1200 days (more than 3 years). The basic assumption underlying use-rate acceleration is that life can be adequately modeled by cycles of operation and cycling rate (or frequency) does not affect the cycles-to-failure distribution.

- Many failures are caused by chemical degradation. The rate of degradation can often be accelerated by testing at higher than usual temperatures.

- Units can be tested at higher than usual levels of stress (like mechanical stress, voltage stress, or pressure). A unit will fail when its *strength* drops below the level of stress to which a unit is subjected. Thus a unit at a high stress will generally fail more rapidly than it would have failed at low stress.

All of these methods require extrapolating from a model that relates the accelerating variable to life. The basis for the extrapolation is, preferably, a physical-chemical model, but often detailed physical knowledge is not available and empirical models, combined with past experience is used instead.

As an example, one of the most commonly used physical/chemical models is the Arrhenius relationship that describes the effect that temperature has on the rate of a simple chemical reaction. This relationship assumes that the failure mechanism can be adequately modeled as a simple one-step chemical or diffusion process. This model can be written as

$$\mathcal{R}(\texttt{temp}) = \gamma_0 \exp\left(\frac{-E_a}{k_\mathrm{B} \times \texttt{temp}\,\mathrm{K}}\right) = \gamma_0 \exp\left(\frac{-E_a \times 11605}{\texttt{temp}\,\mathrm{K}}\right) \tag{1}$$

where $\mathcal{R}$ is the reaction of diffusion rate and $\texttt{temp}\,\mathrm{K} = \texttt{temp}\,{}^\circ\mathrm{C} + 273.15$ is temperature in the absolute Kelvin scale, $k_\mathrm{B} = 8.6171 \times 10^{-5} = 1/11605$ is Boltzmann's constant in electron volts per $^\circ$C, and E_a is the effective activation energy in electron volts (eV). The parameters E_a and γ_0 are product or material and failure mode characteristics. It is important that E_a is not temperature dependent.

Of course, in practical applications acceleration models like the Arrhenius relationship may not be appropriate. In applications, however, more complicated physical/chemical models imply a more complicated time transformation function (e.g., Section 18.3.5 of Meeker and Escobar 1998 and Meeker and LuValle 1995).

6.4 Failure time data

Failure-time data are, by far, the most common type of reliability data. Such data provide failure times for units that failed and running times for the unfailed (censored) units. Figure 9 is a failure data event plot for bearing cages installed in jet engines. The data come from Abernethy, Breneman, Medlin, and Reinman (1983).

Log-location-scale distributions are the most widely used distributions for modeling reliability data. This family of distributions includes the Weibull and lognormal distributions. The log-location-scale distribution cdf is

$$F(t; \mu, \sigma) = \Phi\left[\frac{\log(t) - \mu}{\sigma}\right], \quad t > 0. \tag{2}$$

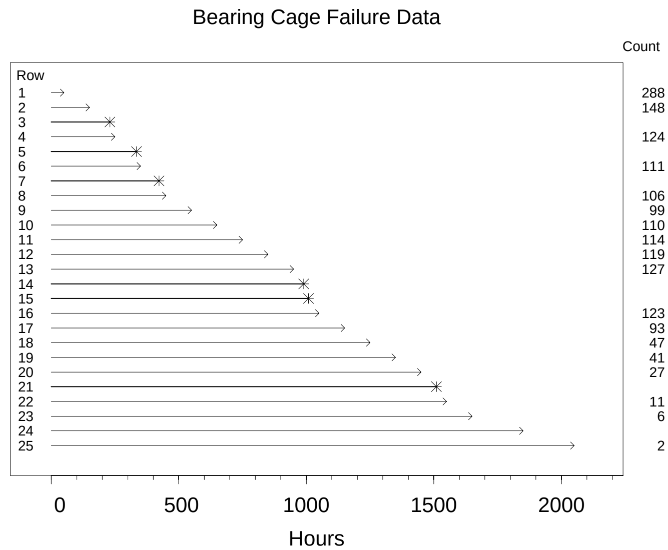

Figure 9: Failure pattern in the bearing cage data

In particular, when $\Phi(z) = \Phi_{\text{sev}}(z) = 1 - \exp[-\exp(z)]$, the standard smallest extreme value cdf, $F(t; \mu, \sigma)$ is the Weibull distribution (with scale parameter $\eta = \exp(\mu)$ and shape parameter $\beta = 1/\sigma$), and when $\Phi(z) = \Phi_{\text{nor}}(z)$, the standard normal cdf, $F(t; \mu, \sigma)$ is a lognormal distribution (with median $\exp(\mu)$ and shape parameter σ). In some cases there is physical/chemical justification for the use of one or the other of these distributions (e.g., see Sections 4.6 and 4.8 of Meeker and Escobar 1998). Many other distributions have also been derived from specific physical/chemical phenomena (e.g., the Birnbaum-Saunders distribution for fatigue-fracture failures). Other distributions have been suggested as more flexible generalizations of existing distributions (e.g., the generalized gamma distribution). Most commonly a distribution is chosen because it fits the data well (or, more appropriately because it has a long history of fitting well for a certain type of failure process).

Figure 10 is a lognormal probability plot of the bearing cage data giving the proportion failing as a function of hours of operation. The plotted points are a modified Kaplan-Meier (KM) estimate (i.e., points are plotted at half of the KM jump height at each failure) and the straight line is the corresponding maximum likelihood (ML) lognormal cdf estimate. The curved line going through the points is the Weibull cdf ML estimate. Lognormal pointwise confidence intervals are also depicted on the plot, expressing statistical uncertainty in the estimates. It is interesting to note that the Weibull ML estimate is approaching the lognormal confidence intervals, reminding us that the confidence intervals do *not* reflect model uncertainty and that extrapolation outside the range of the data can lead to large differences among different models.

Figure 11 shows accelerated life test data on an electronic device (originally from Hooper and Amster (1990), but also analyzed in Chapter 19 of Meeker and Escobar 1998) and the fitted Arrhenius-lognormal model. Combining (1) and (2) with $\Phi(z) = \Phi_{\text{nor}}(z)$ gives the Arrhenius-lognormal accelerated life test model, which can be expressed as

$$\Pr[T(\texttt{temp}) \le t] = \Phi_{\text{nor}}\left[\frac{\log(t) - \mu(x)}{\sigma}\right]$$

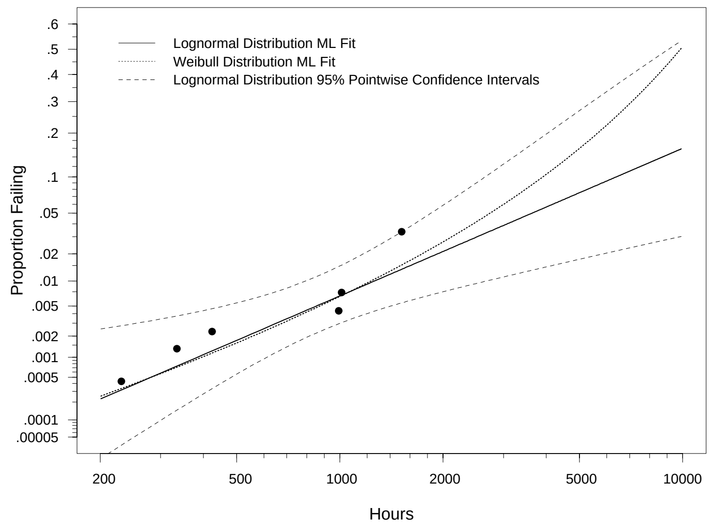

Figure 10: Lognormal probability plot for bearing cage data

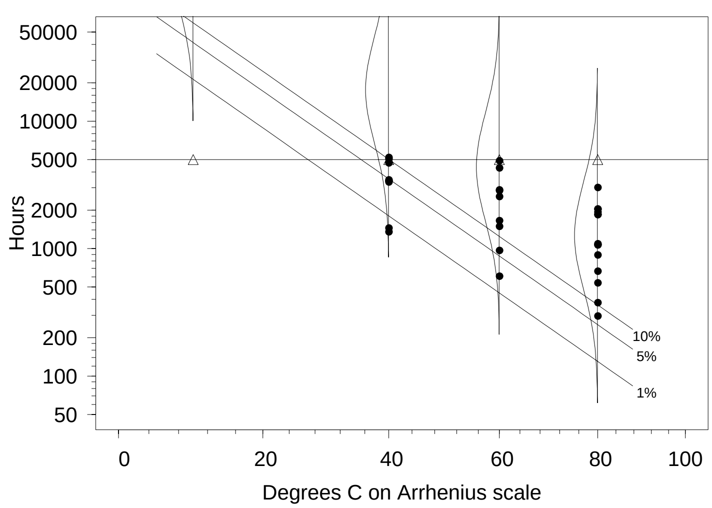

Figure 11: Model plot of the Lognormal-Arrhenius model for accelerated test data for an electronic device

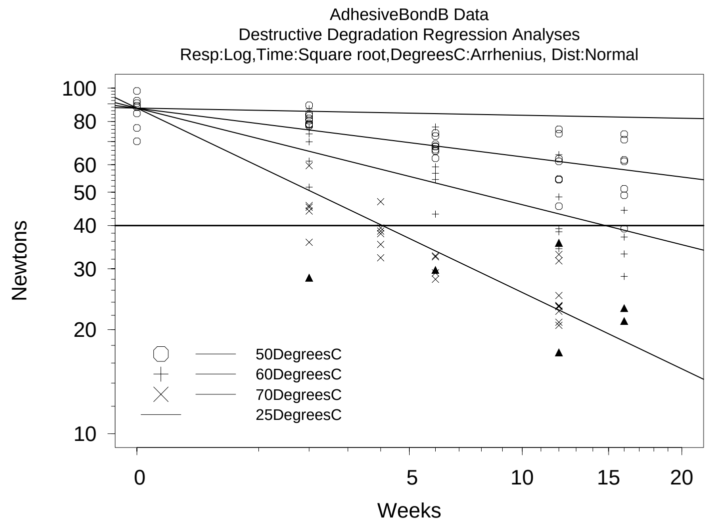

Figure 12: Adhesive bond B accelerated destructive degradation test

where $\mu(x) = \beta_0 + \beta_1 x$, $x = 11605/(\texttt{temp}\,\mathrm{K}) = 11605/(\texttt{temp}\,^\circ\mathrm{C} + 273.15)$, and $\beta_1 = E_a$ is the effective activation energy. The estimates corresponding to the fitted regression model shown in Figure 11 are the ML estimates (ML is needed because of the censoring).

Interestingly the Arrhenius time transformation function implies that σ is constant. This is because the time transformation function is linear and only changes the scale of time.

6.5 Destructive degradation data

Often reliability tests result in few or no failures. In some applications it is possible to measure a degradation variable instead, model the degradation process, and relate degradation to failure. This provides an indirect estimate of a failure-time distribution. Modeling degradation also brings us closer to the physics/chemistry of failure, allowing more information for assessing the adequacy of models.

Figure 12 shows destructive degradation data from an accelerated test on an adhesive used in manufacturing inkjet cartridges. The example comes from Escobar, Meeker, Kugler, and Kramer (2003). Table 1 shows the experimental setup. The 8 units at time 0 were destructively measured for strength (in Newtons) at the start of the experiment, with no aging. The other 80 units were aged at high temperatures and destructively measured at the times indicated in the table.

Note that in Figure 12, the Newtons axis is a logarithmic scale and the Weeks axis is a square root scale. The transformations implied by these scales were initially determined empirically as those that would linearize the degradation versus time relationship at the different levels of temperature, but it was learned later that there is physical/chemical justification for the choices.

The degradation rate is clearly temperature dependent and is assumed to be described by the Arrhenius relationship. This leads to the following model for the degradation strength distribution

Temp	Weeks Aged					
°C	0	2	4	6	12	16
70		6	6	4	9	0
60		6	0	6	6	6
50		8	0	8	8	7
—	8					

Table 1: Adhesive bond B test plan.

as a function of time and temperature

$$F_Y(y;\tau,x) = P(Y \le y;\tau,x) = \Phi\left[\frac{y-\mu(\tau,x,\boldsymbol{\beta})}{\sigma}\right]$$

where $\mu(\tau,x,\boldsymbol{\beta}) = \beta_0 + \beta_1 \exp(\beta_2 x)\,\tau,\ y = \log(\text{Newtons}),\ \tau = \sqrt{\text{Weeks}}$, and $x = -11605/(°\text{C}_j + 273.15)$.

The goal of the experiment was to estimate the failure-time distribution at a nominal use temperature of 25°C. Note the predicted degradation line for 25°C in Figure 12. When strength degrades to 40 Newtons, a unit is defined to be a failure. With this definition, the degradation distribution implies a failure-time distribution which can be derived as

$$\begin{aligned} F_T(t;x,\boldsymbol{\beta}) &= \Pr(T \le t) \\ &= F_Y(\mu_{\text{f}};x,\boldsymbol{\beta}) = \Phi\left[\frac{\mu_{\text{f}}-\mu(\tau,x,\boldsymbol{\beta})}{\sigma}\right] \\ &= \Phi\left(\frac{\tau-\nu}{\varsigma}\right),\ \text{for } t \ge 0 \end{aligned} \tag{3}$$

where

$$\nu = \frac{(\beta_0-\mu_{\text{f}})\exp(-\beta_2 x)}{\mid \beta_1 \mid} \quad \text{and} \quad \varsigma = \frac{\sigma\exp(-\beta_2 x)}{\mid \beta_1 \mid}.$$

and $\mu_{\text{f}} = \log(40)$.

6.6 Repeated measures degradation data

In contrast to the destructive degradation data discussed in Section 6.5, in some applications it is possible to track degradation over time, providing useful information about reliability even if there are no failures. Meeker and Escobar (1998, Chapters 13 and 21) illustrate and describe methods for analyzing repeated measures degradation data.

Figure 13 shows the results of a degradation/life test of a sample of prototype constant-light-output lasers. Over time, a feedback mechanism inside of the laser will increase current to maintain constant light output. Some defective lasers failed catastrophically early in life (indicated by the vertical lines in Figure 13). For the other lasers, after the current increase exceeds 10% of the initial current level, the laser is declared a failure (this is a “soft” failure, as opposed to the catastrophic failures, which are called “hard” failures, because the laser stops working altogether). Assuming that the catastrophic failures could be eliminated by making a change to the product or process design

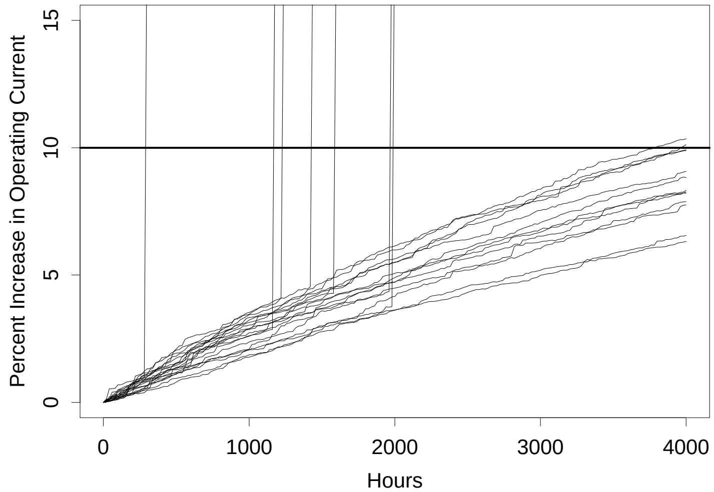

Figure 13: Percent increase in operating current for GaAs lasers tested at 80°C (use conditions 10°C)

(this was indeed done in this application), then a reasonable model for laser degradation would be a linear regression in which the slope is random from unit to unit.

Figure 14 shows repeated measures degradation data from an accelerated test on an solid state RF power amplifier that was to be installed in a satellite system. The experiment and analyses of the data are explained in more detail in Meeker, Escobar, and Lu (1998).

The goal of the test was to estimate the failure-time distribution at 80°C, where the definition of failure was the point in time at which the power drop reached -0.5 dB. The degradation path model fit to the data was derived from the kinetics of the failure-causing chemical reaction. Solving a simple system of differential equations gave

$$\mathcal{D}(t) = \mathcal{D}_{\infty}[1 - \exp(-\mathcal{R}\, t)]$$

where $\mathcal{D}_{\infty}$ and $\mathcal{R}$ were treated at random from unit to unit and the degradation rate $\mathcal{R}$ was modeled as a function of temperature with the Arrhenius relationship. Although it was also possible to estimate the failure-time distribution by using failure-time data, it was felt that the more conservative estimate from the degradation model was more appropriate. This is because the degradation data analysis method used more effectively the information that the units at 150°C were close to failure.

6.7 Recurrence data

Recurrence data arise in a number of important reliability applications, particularly in the analysis of certain kinds of warranty and other field data. Recurrence data arise when a unit (or, more commonly, a group of such units) is monitored over time and a particular event or class of events occurs at points in time. Examples include maintenance actions of repairable systems, system failures, and transactions with customers. Each event may have associated with it a number or numbers describing the event in more detail (e.g., the cost of a maintenance action or the size of an

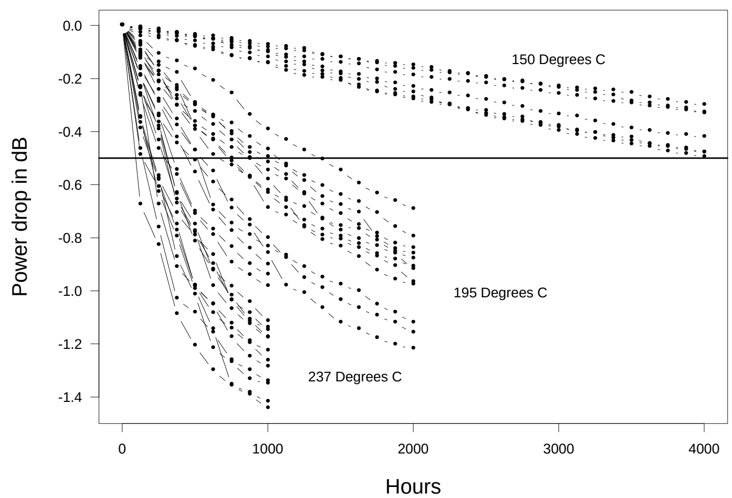

Figure 14: Device-B power drop accelerated degradation test results at 150°C, 195°C, and 237°C (use conditions 80°C)

order). Such data are also known, in the statistical literature, respectively, as point process data and marked point process data. Questions of interest include the recurrence rate (or cost accumulation rate) as a function of time, and whether the rate is increasing, decreasing or constant (again as a function of time). The mean cumulative number of events per unit as a function of time (or mean cumulative cost) is also of primary interest. Such a cumulative functions is called, generically, a mean cumulative function (MCF). Nelson (2003) describes a large number of applications and methods to analyze such data.

As an example, Figure 15 is an event plot for maintenance events for a fleet of earth moving machines. There was a recorded cost associated with each event. Management wanted a model that it could use to predict maintenance costs for future units of the same type that would be operated under similar conditions. The data and analyses are described more fully in Section 16.2.4 of Meeker and Escobar (1998).

Figure 16 is an estimate of the mean cumulative function (MCF), giving an estimate of expected cost per machine, as a function of time. This function can be used directly to answer management's question. The plot also shows 95% approximate confidence intervals for the MCF. The computation of the estimate and corresponding confidence intervals is complicated because of the unequal observation times for the units in service (as seen in Figure 15).

7 Warranty and reliability

Much been written to describe the use of warranty data. The edited books by Blischke and Murthy (1994, 1996) cover a wide range of topics. Robinson and McDonald (1991), Lawless and Kalbfleisch (1992), and Lawless (1998) provide reviews of statistical methods for warranty data.

Warranties are more related to marketing than reliability! In many industries (e.g., the automobile industry), warranty costs, relating to poor quality and reliability, are substantial. Warranty

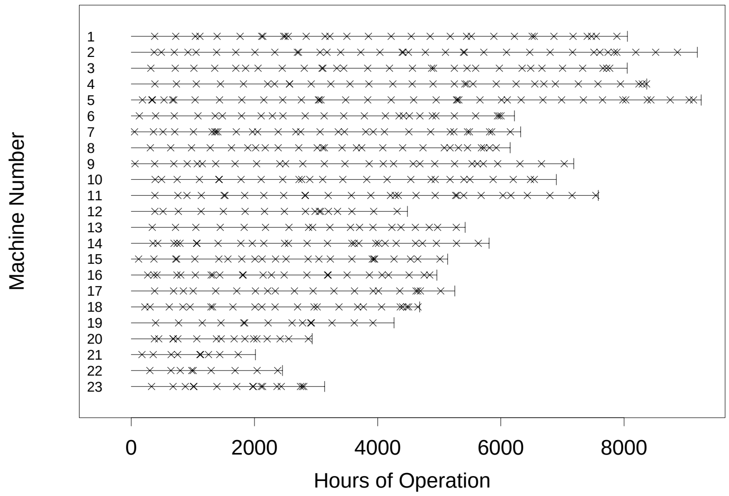

Figure 15: Maintenance events for a fleet of earth-moving machines

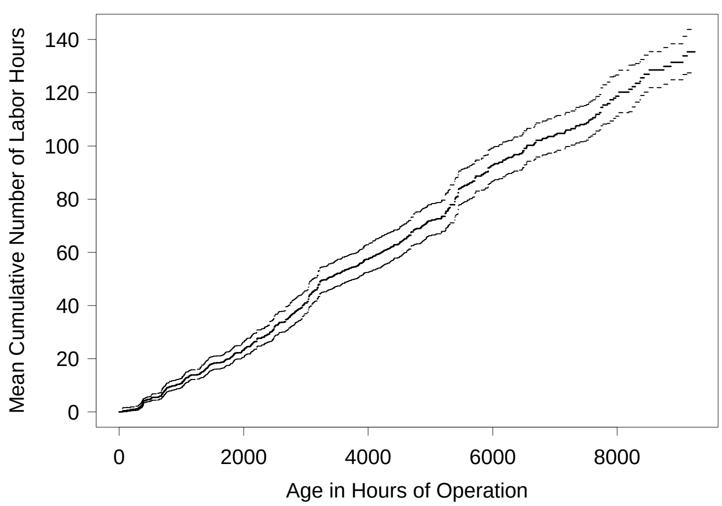

Figure 16: Mean cumulative maintenance cost for a fleet of earth-moving machines

databases exist primarily for required financial reporting. More and more companies are, however, beginning to recognize that warranty data can be useful for:

- Feedback for the design of the next product generation.
- Early warning of unanticipated field problems (see Wu and Meeker 2002).
- Establishing a connection with laboratory testing and environmental characterization.

Warranty data are messy and present serious challenges for proper interpretation and analysis. Warranty data directly reflect one important component of what is seen on a company's bottom line (another important component, which is more difficult to measure, is customers and goodwill lost when a product has a reliability problem).

8 Reliability statistics in practice

As mentioned in Section 1.2, Reliability is, primarily, an Engineering discipline. Nevertheless, Statistics plays an important supporting role within reliability and reliability problems lead to interesting and important statistical problems.

8.1 The role of the statistician on a reliability team

Statisticians play an important role on a reliability team. For example, they

- Contribute to the understanding and modeling of variation.
- Help fill in the gaps in engineering knowledge by designing experiments and interpreting results.
- Help develop systems to ensure that the most meaningful information is obtained to assess reliability and proactively identify problems or potential problems.
- Use appropriate statistical methods to make the most effective use of field and warranty data.
- Develop appropriate methods for combining information from different sources.
- Develop methods for quantifying uncertainty (statistical and model).
- Develop methods (especially graphical methods) for the effective presentation of results.
- Work with other scientists and engineers in the development of appropriate deterministic or stochastic models for physical/chemical failure modes.

8.2 A current example: service life prediction for organic paints and coatings

This section describes a current project involving an interdisciplinary team that is using a modern approach to tackle a difficult problem in reliability. The project is being conducted at the National Institute of Standards and Technology (NIST). More details about the project can be found at http://slp.nist.gov/coatings/cslpmain.html.

The standard method for testing paints and coatings for effects of outdoor exposure is to send specimens to outdoor testing facilities (notably in south Florida and Arizona, for sunny humid and sunny dry conditions, respectively) and to have them exposed to the environment for periods ranging from months to years. Outdoor tests are expensive and take too much time. Manufacturers of paints and coating have been trying for decades, with little success, to develop useful laboratory accelerated testing methods to allow the rapid screening and assessment of the service life of potential new products. The laboratory tests that have been run have tried to mimic outdoor environments by "speeding up the clock" (using increased temperature, increased UV intensity and cycling more rapidly that the usual diurnal cycle). These tests are not reliable and often lead to conclusions that differ from those derived from the data that are returned from the outdoor testing laboratories! Speed-up-the-clock tests do not follow the basic principles of experimental design and thus provide little or no information about the fundamental mechanism leading to degradation and failure.

The NIST approach is to use careful experimentation, measurement, and physical/chemical theory to understand the degradation mechanisms that lead to failure of paints and coatings, starting out with focus on a particular important industrial application. The laboratory experimental setup is based on the NIST SPHERE (Simulated Photodegradation by High Energy Radiant Exposure) technology. This technology, developed by NIST scientists, uses a large integrating sphere to provide a controlled source of UV radiation. There are 32 separate chambers attached to the sphere and within each chamber it is possible to independently control UV intensity, the UV spectrum, temperature, and humidity. Carefully designed experiments will be used to check and refine various mechanistic models for material degradation.

As part of the model development/verification process, outdoor experiments are being conducted at sites in several different climates. At each site specimens are being exposed to actual outdoor conditions with monitoring of UV radiation intensity and spectrum, temperature, and humidity, and the resulting degradation (various physical and chemical measurements are being made for the specimens tested in both the indoor and outdoor exposure tests). Environmental realizations (time series of the environmental variables over time), when used to drive the physical/chemical model, should produce results similar to that seen in outdoor exposure.

Once physical/chemical models degradation are available for a product it will be relatively inexpensive and fast to obtain the information needed to compare different proposed formulations and learn about the effect that different environments will have on reliability.

8.3 Trends in the use of statistics in reliability

The way in which statistical methods are used in reliability has been changing and will continue to do so. Some changes that we can expect to see in the future that will affect the use of statistics in the area of engineering reliability are:

- More up-front problem definition and ensuring that the most meaningful data are obtained during product/process design.
- More use of degradation data and models including stochastic models for degradation.
- Increased use of statistical methods for producing robust product and process designs.
- More use of computer models to reduce reliance on expensive physical experimentation.
- Better understanding of the product environment (e.g., through the use of products outfitted with "smart chips" that record how and in what environment a product is used).
- The availability (through remote monitoring using sensors and modern digital communications) of real-time information on the operational and environmental state of operating systems.

- More efforts to combine data from different sources and other information through the use of "responsible Bayes" methods for combining information from different sources.

8.4 Academic involvement in manufacturing reliability problems

Manufacturing industries have interesting, challenging, technical problems in reliability. There should be more academic involvement in these projects. It would be beneficial for all to have more professors and their students involved in solving these problems. The benefits of increases in such involvement would be:

- The quality of academic research will improve with access to real problems.
- There is a high probability of research impact.
- It can be cost-effective for industry.
- Students and faculty will gain valuable experience.
- It will foster better industry/academic relationships.

It is possible (but sometimes challenging) to achieve these benefits while meeting the requirements of the academic institutions (that research should produce scholarly publications and that external funding is needed to support much of its research).

8.5 Facilitating academic involvement in manufacturing reliability problems

Academics will have to be anxious to get their hands dirty with the difficulties of real problems, including investment of time to learn the language and science of the relevant disciplines. Industrial sponsors (i.e., individuals) will have to invest the time needed to identify, help structure, and provide guidance for selected projects. In today's highly competitive environment, it is difficult for industry to commit time and money unless there is some reasonable expectation of benefits. Some possible approaches that can be used to develop fruitful partnerships between industry and academics include:

- Student internship programs and opportunities for faculty visits to industry (Los Alamos National Laboratory has such a program) that provide assistance to industry and valuable learning experiences for the visitors.
- The NSF GOALI (Grant Opportunities for Academic Liaison with Industry) funding program for academics (students or faculty) to visit industry and/or for industry employees to visit universities to participate in research projects.
- When benefits to industry are uncertain (as will often be the case), industrial funding for cooperative work may be difficult to obtain. In such cases it may be useful for academics to offer to do work for expenses only, at least until a track record has been established.

9 Concluding remarks

Statistical tools that were developed to control and improve quality (e.g., process monitoring and designed experiments) have been useful for improving both quality and reliability. Generally, however, special statistical tools that focus on reliability are needed. With continuing changes in technology, increasing scientific knowledge, and new sources of data and other information, many new interesting problems and challenges lie ahead. Statisticians will continue to have an essential role to play in the area of reliability.

Acknowledgments

We would like to thank Gerald J. Hahn and Angela Jianying Zuo for helpful comments on an earlier version of this paper.

References

ATT (1990) *Reliability by Design*, ATT: Indianapolis, IN.

Abernethy, R. B., Breneman, J. E., Medlin, C. H., and Reinman, G. L. (1983), *Weibull Analysis Handbook*, Air Force Wright Aeronautical Laboratories Technical Report AFWAL-TR-83-2079. Available from the National Technical Information Service, Washington, DC.

Blischke, W. R., and Murthy, D. N. P. (1994), *Warranty Cost Analysis*, New York: Marcel Dekker.

Blischke, W. R., and Murthy, D. N. P. (eds.) (1996), *Product Warranty Handbook*, New York: Marcel Dekker.

Byrne, D., and Quinlan, J. (1993) Robust function for attaining high reliability at low cost, *1993 Proceedings Annual Reliability and Maintainability Symposium*, 183–191, IEEE.

Condra, L. W. (1993), *Reliability Improvement with Design of Experiments*, Second Edition, New York: Marcel Dekker, Inc.

Davis, T. P. (1998), The fallacy of reliability prediction in automotive engineering. *Automotive Excellence*, Spring 1998, 19–21. Published by The American Society for Quality (also available from http://www.timdavis.co.uk/predict.PDF).

Escobar, L. A., Meeker, W. Q., Kugler, D. L., and Kramer, L. L. (2002), Accelerated Destructive Degradation Tests: Data, Models, and Analysis. Preprint 2003-01, Department of Statistics, Iowa State University.

Grove, D. M., and Davis, T.P. (1992), *Engineering, Quality and Experimental Design*, Longman.

Hooper, J. H., and Amster, S. J. (1990), Analysis and presentation of reliability data, in *Handbook of Statistical Methods for Engineers and Scientists*, Harrison M. Wadsworth, Editor. New York: McGraw Hill.

Høyland, A., and Rausand, M. (1994), *System Reliability Theory: Models and Statistics Methods*, John Wiley & Sons.

Kackar, R. N. (1985), Off–Line Quality Control, Parameter Design, and the Taguchi Method (with discussion), *Journal of Quality Technology,* **17**, 176–209.

Lawless, J. F. (1998), "Statistical Analysis of Product Warranty Data," *International Statistical Review*, 66, 41–60.

Lawless, J. F., and Kalbfleisch, J. D. (1992), "Some Issues in the Collection and Analysis of Field Reliability Data," in *Survival Analysis: State of the Art*, Editors J. P. Klein and P. K. Goel, Kluwer Academic Publishers, 141–152.

Logothetis N., and Wynn, H. P. (1994), *Quality Through Design*, Oxford: Clarendon Press.

McLean, H. (2000), *HALT, HASS and HASA Explained: Accelerated Reliability Techniques*, American Society for Quality Press.

Meeker, W. Q., Escobar, L. A., and Lu, C. J. (1998), Accelerated Degradation Tests: Modeling and Analysis. *Technometrics* 40, 89-99.

Meeker, W. Q., and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: John Wiley & Sons.

Meeker, W. Q., and LuValle, M. J. (1995), An Accelerated Life Test Model Based on Reliability Kinetics. *Technometrics* 37, 133-146.

Meeker, W. Q., and Escobar, L. A. (2003), *SPLIDA (S-PLUS Life Data Analysis).* www.public.iastate.edu/~splida

MIL-HDBK-189 (1981), *Reliability Growth Management*. Available from Naval Publications and Forms Center, 5801 Tabor Ave, Philadelphia, PA 19120.

Nelson, W. B. (2003), *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*, (ASA-SIAM Series on Statistics and Applied Probability), ASA-SIAM.

Phadke, M. S. (1989), *Quality Engineering Using Robust Design*, Englewood Cliffs: Prentice Hall.

Robinson, J. A. , and McDonald, G. C. (1991), "Issues Related to Field Reliability and Warranty Data," in *Data Quality Control: Theory and Pragmatics*, Editors Gunar E. Liepins and V. R. R. Uppuluri, Marcel Dekker, Inc, 69–89.

Wu, C. F. J., and Hamada, M. (2000) *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: John Wiley & Sons.

Wu, H., and Meeker, W. Q. (2002), Early Detection of Reliability Problems Using Information from Warranty Data Bases. *Technometrics* 44, 120-133.