RELIABILITY: THE OTHER DIMENSION OF QUALITY

W. J. Youden Memorial Address

by William Q. Meeker
Department of Statistics
Iowa State University

William Q. Meeker

W. J. Youden. It is a pleasure to be here today to give this address in remembrance of the life and the work of W.J. Youden. I was in my third year of college and in my first statistics course in the spring of 1971, when Jack Youden passed away. It was two years later, in a course on experimental design, where I first heard Youden's name. There, I learned about latin square designs and then about Youden square designs. Youden square designs is a term that was coined by R.A. Fisher to describe the results of Youden's novel ideas in the area of designing experiments with balanced incomplete blocks. Jack Youden is probably best known for this and his other work in experimental design. What is less well known is that this early work, like much of Fisher's work, was motivated by experimental work in the agricultural sciences. Jack Youden spent the early years of his career at Boyce Thompson Institute for Plant research in Yonkers, NY. Youden published his balanced incomplete block methods in a 1937 paper with the title "Use of incomplete block replications in estimating tobacco mosaic virus." Part of the message in my presentation today, which is elsewhere supported by similar historical evidence, is that most statistical research with impact and lasting value had its roots in the applied sciences, including engineering.

During his years at the National Bureau of Standards (now the National Institute of Standards and Technology) Jack Youden had a profound effect on the development of statistical methods. Among many other things, he developed and popularized statistical methods for conducting interlaboratory studies. These methods continue to be in wide use around the world. The American Statistical Association has an annual reward, in honor of Jack Youden, in the area of interlaboratory testing.

Background. A substantial amount of the practical experiences that I have had during my career has been in the area of Reliability. I have had the good fortune to have worked with a large number of outstanding industrial statisticians as well as other scientists and engineers.

During the three summers when I was a graduate student, I had an internship at General Electric Corporate Research and Development Center (Now General Electric Global Research Center), working on projects under the supervision of Gerry Hahn and Wayne Nelson. I have continued to benefit over the years from collaboration and interaction with Gerry, Wayne, Necip Doganaksoy, Roger Hoerl, and others at GE. From 1978 to 1992, I spent the better part of each summer working for the quality/ reliability organization (the actual name of the department changed a number of times over this period of time) at Bell Laboratories, learning first hand about statistics, quality and reliability from such experts as Blan Godfrey, Jeff Hooper, Ramón León, Mike Tortorella, Vijay Nair, Michele Boulanger, Mike LuValle and various other scientists and engineers too numerous to mention here. My presentation today, based on these and other experiences, will describe the relationship between quality engineering, reliability engineering, and statistics. There will be some links to the life and work of Jack Youden.

Reliability. 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.

The usual textbook definition of reliability reads something like "the probability that a unit will survive until a given point in time under specified use conditions." A more appropriate definition is "the probability that a unit will survive 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 (2001) states "Reliability is product performance 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

Continued on page 5

Continued from page 4

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.

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 product engineering and 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."

Robustness. Robustness 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 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 operation and to do this in a manner that is economically efficient.

The operational/technical ideas behind robustness derive from the important engineering ideas that were brought to us by Genichi Taguchi. Taguchi suggested a strategy, based on ideas of statistically designed experiments that can be used to improve product or process design 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 the book-length treatments by Phadke (1989), Grove and Davis (1992), Wu and Hamada (2000), and Condra (2001).

Engineering design can be viewed as a complicated optimization problem (although I am not sure how useful it is to do so, given our inability to quantify all of the inputs and especially intangible costs). Given this view, I have often wondered, in conversation with statisticians who work in industry, whether the tools of robustness would be useful for a product or process design that already had been very well engineered (and thus was already very close to optimum). I have been assured, almost uniformly, that few, if any, such designs exist, although clearly this must be product and industry specific.

Quality and reliability. Figures 1 through 4 illustrate one view of the relationship between quality and reliability. Suppose that the distributions represent some arbitrary 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 or outside either of the specification limits. Figure 1 reflects barely acceptable 3-sigma quality.

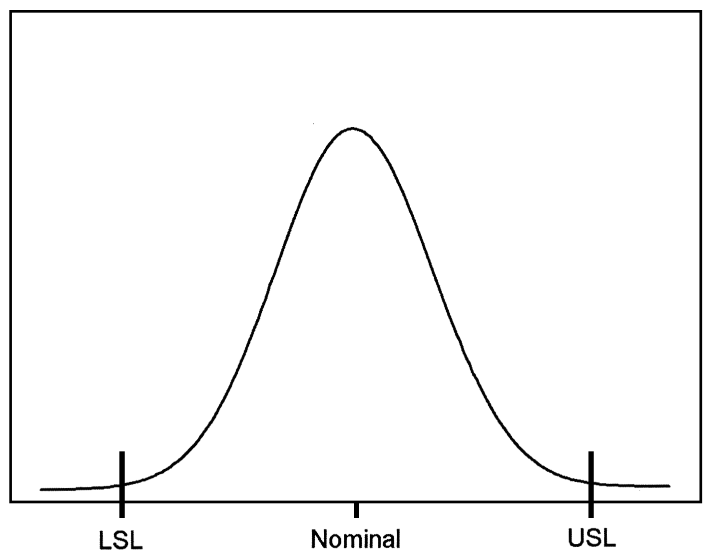

Figure 1. Three-sigma quality

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 2, 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.

Continued on page 6

W.J. YOUDEN MEMORIAL ADDRESS

Continued from page 5

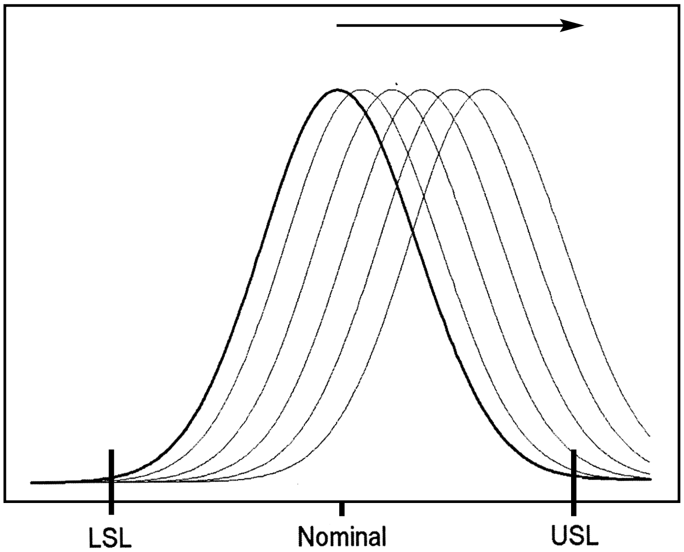

Figure 2. Drifting three-sigma quality

Figure 3 contrasts good quality and poor quality.

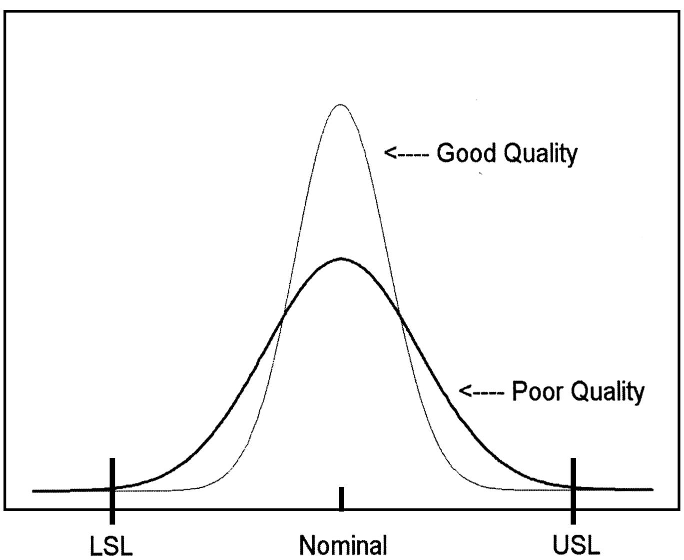

Figure 3. Good and bad quality

As illustrated in Figure 4, with good quality, even with expected drift over time, customers will, generally, continue to have good performance---quality over time or high reliability.

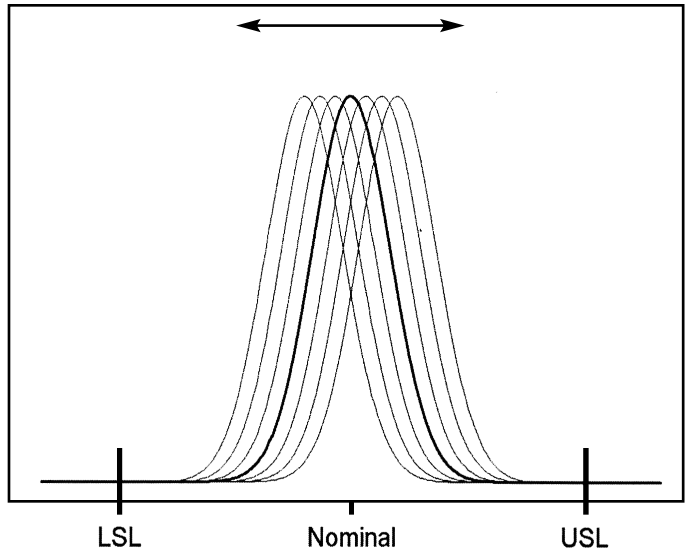

Figure 4. Drifting six-sigma quality

As with quality, we see that variability is the enemy of reliability. Examples of important sources of variability are manufacturing (including raw materials), environmental conditions (e.g., stresses), and customer use rates. Reduction of input variability and reduction in the transferal of input variability to customer perceivable variability are important goals for engineering design.

Reliability demonstration versus reliability assurance. Managers, for example, need information about reliability before making product-release decisions. Potential customers need information reliability before deciding to purchase a product.

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 less than the target.” For example, using minimal assumptions (including a go no-go assessment), to demonstrate that reliability at time t_0 hours is 0.99, with 90% confidence, requires testing at least 230 units for t_0 hours with zero failures. To have just an 80% chance of passing the test requires that the true reliability be approximately .999! The required sample size can be reduced, to some degree, by making certain assumptions (e.g., about the form of the failure-time distribution).

It might be feasible to run such large tests for a material or a low-level component. For a complicated, expensive systems or subsystem, however, such traditional reliability demonstration tests are not practicable. Reliability assurance is the alternative.

Continued on page 7

W.J. YOUDEN MEMORIAL ADDRESS

Continued from page 6

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.

Structured programs for design for reliability. The concept of reliability demonstration is not new. 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 upfront 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.

Failure modes and reliability prediction. 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. 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). Usually it is the unanticipated failure modes that cause the most serious reliability problems. It is for this reason that an important component 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 engineering knowledge and previous experience, failure modes and effects analysis (FMEA) in up-front design, highly accelerated life testing (known as "HALT" tests wherein prototype subassembly units or systems are subjected to higher than usual operating/environmental conditions in order to shake-out design weaknesses), and early feedback from the field. Some products undergo "beta testing," where early-production units of a product are released for use, preferably in a high-use, customer-friendly 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.

In many applications, engineers try to use a predictive model that will provide at least a rough idea of a product's reliability in the field. Although widely practiced, the development and use of reliability models is 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 failures is that they are important parts of the learning process for an important task.

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.

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 stress and the effect of such "spectra" of stresses needs to be part of the reliability model.

Determining the needed inputs under stringent time and economic constraints is always a challenge. Typical sources of information (roughly in order of 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.

Continued on page 8

W.J. YOUDEN MEMORIAL ADDRESS

Continued from page 7

When none of the usual sources of information provide the needed information to the desired degree of accuracy, engineers will typically build in 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." I 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.

Distinguishing features of reliability 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 (mean and variance) and ordinary least squares rarely provide appropriate methods of analysis (instead simple graphical methods combined with maximum likelihood estimation for fitting parametric models are used)
- Model parameters and regression coefficients are not of primary interest (instead, failure rates, quantiles, probabilities are needed)
- Extrapolation 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).

Reliability data.

Traditionally, most reliability data was failure time data, reporting the failure times for units that failed and the running times for the unfailed (censored) units. In many reliability studies of high reliability products, few or no failures are observed. 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) illustrate and describe methods for analyzing both kinds of data. In other applications a sequence of recurrent events are observed on a sample of units (e.g., system repair or maintenance actions). Special "recurrence data analysis" methods have been developed for such data (see Nelson 2003).

U**se 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.

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

- Feedback for the design of the next product generation
- Early warning of unanticipated field problems
- 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).

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.

Continued on page 9

Continued from page 8

A current example: service life prediction for organic paints and coatings. I thought that it would be useful to describe a current project where an interdisciplinary team 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). I have been serving as an advisor to the project.

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 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 accelerated testing methods that allow the rapid screening and assessment of the service life of potential new products. The 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 than 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.

Jon Martin with the Materials and Construction Research Division, Building and Fire Research Laboratory at NIST is the project leader. Jon is leading a NIST consortium with a number of industrial partners and his NIST scientific staff (which includes chemists, materials scientists, physicists, and engineers). More details about the project can be found at http://slp.nist.gov/coatings/cslpmain.html.

The approach is to use careful experimentation 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 model degradations are available for a product, it will be relatively inexpensive to obtain information needed to compare different proposed formulations and learn about the effect that different environments will have on reliability.

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 the most meaningful data are obtained during product/process design.
- More use of degradation data and models including stochastic models
- 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")
- 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.

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 such involvement would be

- The quality of academic research will improve with access to real problems
- High probability of research impact

Continued on page 10

W.J. YOUDEN MEMORIAL ADDRESS

Continued from page 9

- Cost-effective for industry
- Students and faculty will gain valuable experience
- 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).

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 internships and opportunities for faculty visits to industry (Los Alamos National Laboratory has such a program) provide assistance to industry and valuable learning experiences for the visitors.
- The NSF GOALI (Grant Opportunities for Academic Liaison with Industry) program provides funding 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 to academics to offer to do work for expenses only, at least until a track record has been established.

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.

Acknowledgements. I would like to thank Luis A. Escobar and Gerald J. Hahn for helpful comments on the slides for the original talk and this written version.

References

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

Condra, L.W. (2001), *Reliability Improvement with Design of Experiments*, Second Edition, Marcel Dekker.

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

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

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*, Prentice Hall.

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