
Trends in the Statistical Assessment of Reliability

William Q. Meeker

Department of Statistics and Center for Nondestructive Evaluation, Iowa State University,
Ames, Iowa 50010 wqmeeker@iastate.edu

Summary. Changes in technology have had and will continue to have a strong effect on changes in the area of statistical assessment of reliability data. These changes include higher levels of integration in electronics, improvements in measurement technology and the deployment of sensors and smart chips into more products, dramatically improved computing power and storage technology, and the development of new, powerful statistical methods for graphics, inference, and experimental design and reliability test planning. This paper traces some of the history of the development of statistical methods for reliability assessment and makes some predictions about the future.

Keywords and phrases: Accelerated testing, Bayesian methods, Degradation data, Maximum likelihood, Multiple failure modes, Nonparametric estimation, Product design, Recurrence data, Statistical software, Warranty data

1 Background and purpose

Today's manufacturers need to develop new, higher technology products in record time while improving quality, reliability, and productivity. Much of this need has been driven by the expanding global marketplace and the resulting increased competition. Manufacturers of high quality and high reliability products have a strong competitive advantage. These manufacturers are, however, facing:

- Need for rapid product development
- Changing technologies/new materials
- More complicated products with more components
- Higher customer expectations for better reliability

As suggested by Condra [Con93], reliability can be defined as *quality over time*. The improvements in quality of manufactured products that we have seen over the past 30 years (e.g., in automobiles manufactured in the United States) have also had the effect of improving product reliability. It has been recognized, however, that achieving and improving high reliability requires tools that lie beyond the standard tools used in quality improvement. Generally achieving high reliability requires careful focus on the time dimension.

Reliability is a highly-quantitative engineering discipline. Probability theory and statistical models and methods play an important role in reliability. In particular, many reliability-related decisions require the analysis of reliability data, either from past field performance or from laboratory tests (usually accelerated tests). One important consideration, that separates reliability from many other applications of statistics, is that extrapolation is present in almost all applications. For example, we extrapolate in time when we have one year of data but have to predict warranty returns going out three years. We extrapolate in temperature when we test units at high temperatures and then estimate life at use conditions. We extrapolate from past experience when we use knowledge about a model that worked well in the past to describe a new situation. After conducting a laboratory test, we extrapolate to behavior in the field.

The purposes of this non-technical paper are to outline some trends in the use of statistics in reliability, connect these with changes in technology, and to predict what we can expect to see in the future, indicating areas where more research will be needed.

2 Traditional reliability data and data analysis

Reliability data arise from a number of different sources, including laboratory life tests, field tracking studies, and warranty data bases. Traditional reliability data have consisted of failure times for units that failed and running times for units that had not failed. Interestingly (as can be seen by reading old papers in the engineering literature), there was a long period of time when many engineers thought that it was necessary that all units fail so that the life data could be analyzed. Of course, methods for computing estimates of failure distributions from what is essentially censored data have been used for centuries in medical and insurance applications. Methods for analyzing censored data (nonparametric estimation and maximum likelihood) were further developed in the 1950s and the 1960s and became well-known to most statisticians by the 1970s. The bibliographic guide provided by Buckland [Buc64] nicely outlines the important references in this area up to that point in time. Books written by Mann, Schafer, and Singpurwalla [MSS74], Lawless ([Law82] second edition [Law03]), Nelson [Nel82], Cox and Oakes [CO84], and Crowder et al. [CKSS91], document many of the important advances made during the 1960s and 1970s. Statistical methods for the analysis of censored data began to appear in commercial software by the mid-1980s (starting with SAS) and are commonplace today.

The most popular tool for life data analysis is the probability plot, used to assess distributional goodness of fit, detect data anomalies, and to display the results of fitting parametric distributions in the presence of censored data. It should be more widely recognized, however, that a probability plot is a valuable tool for general data analysis, even when there is not any censoring. A community of engineers has long championed what has been called *Weibull analysis* which implies fitting a Weibull distribution to failure data. Weibull analysis, for example, was described in Abernethy, Breneman, Medlin and Reinman, [ABMR83]. The Weibull distribution is not, however, always the best distribution to use, and modern software allows fitting a number of different parametric distributions. The vast majority of applications in reliability use either the Weibull or lognormal distribution. One reason for this is that there are strong mechanistic justifications that suggest these distributions, much as the central

limit theorem can sometimes be used to explain why some random variables should be well-described by a normal distribution.

The method of maximum likelihood (ML) and likelihood-based inference methods are at the core of almost all reliability applications. For example, all of the statistical methods (even the nonparametric methods) used in Meeker and Escobar [ME98] are based on ML. There are two primary reasons for using ML. First, ML is highly versatile; it is hard to find statistical problems where ML cannot be used. Secondly, under mild conditions, ML methods are known to be optimum in large samples. Even when these mild conditions do not hold, it is hard to find competitors that are consistently better.

In most applications of reliability, it is important to quantify statistical uncertainty (i.e., uncertainty due to limited data). Confidence intervals are the most commonly used method to do this. Normal-theory based large-sample approximate confidence intervals are still the most commonly used method to construct confidence intervals, in spite of the fact that there have been examples and studies to show that these intervals can be horribly inadequate (i.e., having coverage probability from the nominal value), even with large sample sizes. Either likelihood-based confidence intervals or simulation-based confidence intervals dominate the crude normal approximations (e.g., Meeker [Mee87], Vander Weil and Meeker [VM90], and Jeng and Meeker [JM00]). Before too long, improving computer power and clever implementation schemes should, for many applications, eliminate the use of normal-approximation intervals.

3 Product design and reliability budgeting

The use of probabilistic modeling for reliability is becoming more common in product design. Historically, engineers used deterministic models for product design and applied *factors of safety* to assure that a product would be reliable. In a highly competitive market, however, being overly conservative may require a price that is too high for the market. In recent years, many companies have instituted *Reliability by Design* or *Design for Six Sigma* programs that involve some kind of reliability assurance process based on probabilistic system reliability models. Such models are described in books on system reliability such as Rausand and Hoyland [RH04]. These models require, as inputs, the reliability of individual components and interfaces. Engineers need to obtain component reliability information in a timely and economical manner. The most common and least expensive sources of such information are standard values in handbooks (e.g., MIL-HDBK-217F [MIL91]) and previous field experience. If the information is not available from these relatively inexpensive sources (e.g., because the component is new or an old component is being used in a different application or environment that could affect its reliability), engineers may have to conduct their own tests in the laboratory. Most such tests would be *accelerated* so that the information is obtained in a timely manner.

Another source of component life information that is becoming more widely used is physical/chemical or other mechanistic models for failure. For example, finite element models can sometimes be combined with other engineering/physical knowledge to predict reliability of a component. Such methods are described, for example, in Haldrar and Mahadevan [HM00]. Of course, in many applications, some actual physical testing is still required to provide inputs to the models or to verify parts of the models. The use of such models in reliability often requires large amounts of time and effort in

development. The hope is that such investments will result in lower costs in the future due to savings from not having to do as much actual physical testing.

4 Accelerated testing

When a component must last for years or even decades, it may be possible to conduct an *accelerated test* to obtain information in a timely manner. Virtually all life tests done to evaluate reliability are accelerated in one way or another. The basic idea is to test units at high levels of cycling rate, temperature, voltage, stress, or another accelerating variable to get reliability information quickly. Then a physically-motivated model is used to extrapolate to use conditions. Nelson [Nel04], originally published in 1990, is the most important reference on the statistical aspects of accelerated testing. This book describes models, statistical methods for analysis, and statistical methods for test planning.

It is sometimes said, jokingly, that engineers are very good at conducting accelerated tests that cause units to fail quickly. Hitting a sample of integrated circuits with a hammer will certainly make them fail, but will provide no useful information about product reliability. The important, serious, question is whether the failures generated in the accelerated test provide useful information about how the product will behave at actual use conditions. There are numerous examples (e.g., two of the three examples in Chapter 19 of Meeker and Escobar [ME98]) where using too much stress or a temperature that was too high caused a new failure mode that, if not recognized and accounted for, would lead to overly optimistic estimates of life at use conditions. Thus, the most important challenge of accelerated testing is deciding how to accelerate, how much to accelerate, and to find an adequate model to relate the results from the accelerated test to actual use conditions.

5 Multiple failure modes

For some applications, when analyzing reliability data, it is important to distinguish among different product failure modes. The reasons for this are:

- When the failure modes behave differently (e.g., some are defect or infant mortality related and others are caused by wearout), it is generally easier to find a well-fitting failure-time distributions for the individual failure modes (e.g., Example 15.6 in Meeker and Escobar [ME98]).
- When forecasting warranty costs, some failure modes are much more expensive to fix than others (e.g., replacing a mother board versus replacing a defective battery in a computer). In some applications, there is one failure mode that is of critical importance (e.g., a failure mode that could cause serious harm) and others that are innocuous, leading to end of life of the product and thus eliminating the possibility of the critical failure mode.
- Predictions are often needed for the number of replacement parts that will be needed to effect repairs.
- Knowledge of the relative frequency of different failure mode and the effect of eliminating one or more of the individual failure modes is important for engineers who

need to make design changes that will improve product reliability and reduce warranty costs.

When failure mode information is available for all failed units and when the different failure modes can be assumed to be statistically independent, the analysis of multiple failure mode data is, technically, not much more difficult than it is for analyzing a single failure mode.

Competing risk theory provides the appropriate statistical model for life data with multiple failure modes. The first book-length treatment of the theory of competing risk was provided by David and Moeschberger [DM78]. For single distribution applications in reliability, these methods are described and illustrated with examples in Chapter 5 of Nelson [Nel82], Chapter 15 of Meeker and Escobar [ME98], and Crowder [Cro01]. Nelson [Nel90] describes methods for analyzing accelerated life test data with multiple failure modes. Life data analysis with multiple failure modes is, however, greatly facilitated when software tools have been designed to make the needed operations easy. Today, several statistical packages provide capabilities for estimating separate distributions for each failure mode and for making assessments of improvement in product reliability by eliminating one or more of the failure modes.

When failure-mode information is missing for some or all of the failed units (known as masked causes of failure) or when the failure modes cannot be described by the simple independence model, the analysis is more complicated and special methods, not generally available in software, have to be employed. Flehinger, Reiser, and Yashchin [FRY98], for example, describe statistical methods for dealing with masked data.

When the failure times for different failure modes are not independent, one approach is to use some kind of multivariate failure-time model. In practice, however, there is not enough information to identify such a model. When it is important to estimate the marginal distributions for each failure mode (e.g., to predict the effect of removing a failure mode), it is often possible to collapse a large number of related failure modes into several groups such that the failure modes are approximately independent. Such collapsing of failure modes is common in practice, (e.g., page 34 of David and Moeschberger [DM78]). Meeker, Escobar, and Hong [MEH09] describe an application where many failure modes were collapsed into two groups, engineering judgment was used to make an assumption about the degree of dependency, and sensitivity analysis was used to assure that inferences were reasonably accurate over a plausible range of assumptions. In some applications, estimation of the sub-distribution functions (e.g., page 4 of Crowder [Cro01]), which are identifiable, is sufficient.

6 Field and warranty data

Although laboratory reliability testing is often used to make product design decisions, the real reliability data comes from the field, often in the form of warranty returns or, specially-designed field-tracking studies. Warranty databases were initially created for financial-reporting purposes, but more and more companies are finding that warranty data is a rich source of reliability information. Perhaps six to eight months after a product has been introduced into the market (sooner if costs have already been higher than expected), managers begin to ask about warranty costs over the life-cycle of the product. Two common problems with warranty data are that good failure mode

information is rarely available (there is usually some kind of code in the database, but it is usually of limited use to determine the actual cause of failure) and warranty data are generally heavily censored, because information that is obtained after a product is out of warranty is limited. Thus, even though companies should be concerned about reliability of their products far beyond the end of a warranty period, operationally, little data is available.

For some products, careful field tracking provides good reliability data. Examples include medical devices and a company's fleet of assets (e.g., information about the reliability for a fleet of automobiles).

There are sometimes large gaps between predictions made from product design models (supplemented by limited reliability testing) and reality. These differences are often caused by unanticipated failure modes. Algorithms for early detection of emerging reliability issues (e.g., Wu and Meeker [WM02]) are being implemented in software and have the potential to save companies large amounts of money. Once a new emerging issue has been identified, statistical methods (e.g., Escobar and Meeker [EM99], and Lawless and Fredette [LF05]) can be used to produce forecasts for the additional warranty costs. Field data also provides important feedback that can be used to assess the adequacy of reliability predictions methods and to give engineers information on how to design future products.

7 Degradation reliability data

In modern high-reliability applications, we might not expect to see failures in our reliability testing, resulting in limited information about reliability needed for product design. Suppose that 500 units had been put on test and run for 1000 hours of operation for a product that is required to last 5000 hours. If there are no failures at the end of the test, there would be little or no information for quantifying reliability (depending on assumptions that one might be willing to make). If, however, we could monitor, over time, a degradation (or a performance) variable that is closely related to failure (e.g., length of a fatigue crack or light output from a laser) on all of the test units, there would be a large amount of reliability information.

There are a number of other advantages for using such repeated-measures degradation data for reliability assessment. For example, degradation data provides information that is much richer for building and assessing the adequacy of physical/chemical models used for test acceleration.

Today the term *degradation* refers to either performance degradation (e.g. light output from an LED) or some measure of actual chemical degradation (e.g., concentration of a harmful chemical compound). Over the past 30 years, we have seen many different kinds of applications where degradation data were available. Some early ideas on the use of degradation models in reliability were given in Gertsbakh and Kordonsky [GK69]. Through the 1990s and continuing today, statistical methods have been developed for making reliability inferences from degradation data. Initially these were developed by researchers or engineers in need of the methods. Statistical methods for the analysis of degradation data are, however, now beginning to be deployed in commercial statistical software. Some engineers (e.g., Murray [Mur93]) had been using informal *simple* methods of analysis that fit models to the sample path for individual units and extrapolated these until some failure level, providing *pseudo failure* data that could be analyzed by

common life data analysis methods. Lu and Meeker [LM93] and Meeker, Escobar, and Lu [MEL98] used a more sophisticated random effects model to describe unit-to-unit variability and showed how the degradation model, along with a failure definition, induce a failure time distribution. In some areas of application, it is necessary to model the stochastic behavior in the sample paths over time. Lawless and Crowder [LC04], for example use such a model.

There have been a number of examples where the natural response in a reliability test is a degradation variable, but the analysts (at least initially) turned the degradation data into failure-time data, because all of the test books and software known to them dealt only with the analysis life data. The application described in Meeker, Escobar, and Lu [MEL98] is one such application. In these examples, the limited number of failures provided only limited reliability information and the results of a degradation analysis was more informative.

When an appropriate degradation variable can be measured, degradation data, when properly analyzed, can provide much more information because there are quantitative measurements on all units (not just those that failed). Indeed, it is possible to make powerful reliability inferences from degradation data even when there are no failures. It is, of course, not always possible to find a degradation variable that corresponds to a failure mode of concern.

Even when a repeated measures degradation variable is not available, it might be possible to do destructive tests to evaluate units that have not failed. Examples of such *destructive degradation tests* are given in Chapter 11 of Nelson [Nel90] and in Escobar, Meeker, Kugler, and Kramer [EMKK03].

8 Recurrence data

The discussion in the previous sections dealt with reliability data analysis for nonrepairable components (or devices). Since a nonrepairable component can fail only once, time to failure data from a sample of nonrepairable components consist of the times to first failure for each component. In many applications involving nonrepairable components, the assumption of independent and identically distributed failure times (or at least deviations from an assumed model that might have more than one component of variability) is a reasonable one and suitable lifetime distributions (such as the Weibull or lognormal) are used to describe the distribution of failure times. In contrast, repairable system data typically consist of the times of multiple repairs (or other events of interest) on the same system. Such data are known as recurrence data. In the discussion here we will, for concreteness, refer to recurrences as *repairs*. The purpose of some reliability studies is to describe the trends and patterns of repairs of failures for an overall system or collection of systems, over time. Data consist of a sequence of system repair times for similar systems. When a single component or subsystem in a larger system is repaired or replaced after a failure, the distribution of the time to the next system repair will depend on the nature of the repair and the overall state of the system at the time just before the current repair and the nature of the repair. Thus, repairable system data, in many situations, should be described with models that allow for changes in the state of the system over time or for dependencies between repairs over time. A number of books have been written that describe the many technical advances in this area over the past 20 years. Nelson [Nel03] describes basic graphical and simple,

but effective, nonparametric methods for recurrence data analysis and illustrates the methods with a wide range of applications. These methods have been implemented in several computer packages. Cook and Lawless [CL07], which was written at a higher technical level, describes methods and applications involving a wide variety of nonparametric, parametric, and semiparametric recurrence data models, including regression models. The authors show how readily available software can be used to implement the methods described in their book.

9 The next generation of reliability data

Due to changes in technology, the next generation of reliability field data will be richer in information. Use rates and environmental conditions are important sources of variability in product lifetimes. The most important differences between carefully controlled laboratory accelerated test experiments and field reliability results are due to uncontrolled field variation (unit-to-unit and temporal) in variables like use rate, load, vibration, temperature, humidity, UV intensity, and UV spectrum. Historically, use rate/environmental data has, in most applications, not been available to reliability analysts. Incorporating use rate/environmental data into our analyses will provide stronger statistical methods. Today it is possible to install sensors and smart chips in a product to measure and record use rate/environmental data over the life of the product. In addition to the time series use rate/environmental data, we also can expect to see further developments in sensors that will provide information, at the same rate, on degradation or indicators of eminent failure. Depending on the application, such information is also called *system health* and *materials state* information. In some applications (e.g., aircraft engines and power distribution transformers), system health/use rate/environmental data from a fleet of products in the field can be returned in real time to a central location for real-time process monitoring and especially for prognostic purposes. An appropriate signal in these data might provoke rapid action to avoid a serious system failure (e.g., by reducing the load on an unhealthy transformer). Also, should some issue relating to system health arise at a later date, it would be possible to sort through historical data that have been collected to see if there might have been a detectable signal that could be used in the future to provide an early warning of the problem.

In products that are attached to the Internet (e.g., computers and high-end printers), such use rate/environmental data can, with the owner's permission, be downloaded periodically. In some cases, use/environmental data will be available on units only when they are returned for repair (although monitoring at least a sample of units to get information on un-failed units would be statistically important).

The future possibilities for using use rate/environmental data in reliability applications are unbounded. Lifetime models that use rate/environmental data have the potential to explain much more variability in field data than has been possible before. The information can also be used to predict the future environment lifetimes of individual units. This knowledge can, in turn provide more precise estimates of the life time of individual products. As the cost of technology drops, cost-benefit ratios will decrease, and applications will spread.

10 Software for statistical analysis of reliability data

Newly developed statistical methods will not see wide-spread use until they are implemented in easy-to-use software. Thus, it is important to have such software to do reliability data analysis. The first major software system for reliability data analysis was designed and developed by Wayne Nelson. This package, called STATPAC, was described in Nelson and Hendrickson [NH72] and Strauss [Str80]. STATPAC was far ahead of its time and contained a combination of capabilities for graphical analysis and fitting of general statistical models with censored data that are available today in only the most advanced statistical packages. Meeker and Duke [MD81] provided a useful computer program for doing reliability data analysis with fewer capabilities. By the mid-1980s, SAS had incorporated some of the most important models and methods described in Lawless [Law82], Nelson [Nel82], and Nelson [Nel90] into their general purpose system and these capabilities continue to be developed and extended. More recently, JMP, MINITAB, R, and S-PLUS have also incorporated some of the most widely used methods and models for reliability data analysis into their packages. The following list describes some popular packages commonly found on desk-top computers.

JMP(www.jmp.com) is a popular, highly-sophisticated general-purpose desktop statistical software package. In addition to such standard statistical tools such as single distribution analysis, regression analysis, and tools for experimental design, JMP also has some special tools for reliability data analysis, including the analysis of censored data, competing risk analysis, accelerated life data, and the analysis of recurrence data. JMP has appealing tools for graphical presentation of data and fitted models. JMP also has a sophisticated scripting language that allows extension of the system.

MINITAB (www.minitab.com) is another popular, general-purpose desktop statistical software package. Its reliability analysis capabilities are similar to those of JMP.

The Reliasoft (www.reliasoft.com) suite of programs does not provide general-purpose statistical capabilities, but rather attempts to cover a broad range of needs of a reliability engineer. WEIBULL ++ does basic analysis of single-distribution data. ALTMA can be used to analyze accelerated life test data. BLOCKSIM provides predictions of system reliability, based on evaluation of a system specified by a description of system structure and the reliability of individual components. RG can be used to assess reliability growth of a system.

S-PLUS (www.insightful.com) is a general-purpose, highly-sophisticated environment for doing graphics and statistical computing, using the S language (which was developed at Bell Laboratories). One of the important features of S-PLUS is that it is extendable. That is, users can add capabilities (including GUIs) at the same level as the developers of the system. SPLIDA (www.public.iastate.edu/splida) is a free add-on to S-PLUS that has extensive capabilities for planning reliability studies and for analyzing reliability data. Almost all of the SPLIDA capabilities are available through the SPLIDA GUI.

R (www.Rproject.org) is a freeware implementation of the S language having many of the same capabilities of S-PLUS, but with only limited GUI capabilities. It is expected that there will soon be an R version of SPLIDA, but because of the limited GUI capabilities in R, this version of SPLIDA will require the use of S-language commands to operate.

11 Use of Bayesian methods in reliability

The biggest changes in the future of reliability analysis will come in the form of much more widespread use of Bayesian methods.

There are, in the folklore of reliability, stories of how the use of Bayesian methods in reliability applications has led to unreasonably optimistic predictions for products that had poor reliability. Combined with healthy skepticism about the use of prior information to predict reliability and the previously difficult technical challenges to implement Bayesian methods, applications of Bayesian methods in Reliability has been limited. Of course, there are also many well-documented stories about serious reliability disasters where Bayesian methods played no role in the faulty decision making.

As mentioned in Section 2, ML is at the core of most statistical methods used for reliability data analysis. Ignoring the philosophical differences in the theory of statistical inference, Bayesian methods can be viewed as an extension of likelihood-based methods, where one combines prior information, in the form of a prior probability distribution, with the likelihood. Bayesian methods have the same appealing characteristics as ML (versatile and good statistical properties), but also have a very important advantage. In particular, Bayesian methods provide a formal method for combining data from various different sources, such as information from different studies, information from reliability tests at different levels of product and component integration, prior information based on previous experience, or prior information from general, but imprecise, engineering knowledge.

Martz and Waller [MW82] provided an early description of the use of Bayesian methods in reliability applications. The applications were, however, limited because of limitations in the technology (both statistical and computing power). Over the past 25 years, however, there has been an explosion in interest and application of Bayesian methods in a wide range of areas of application. This enthusiasm has been driven by important advances in methods for implementing the Bayesian paradigm (especially MCMC methods) and important advances in computer hardware capabilities. Ibrahim, Chen, and Sinha [ICS01] is an important reference on the analysis of survival data, taking applications from the areas of medicine and public health. Hamada, Wilson, Reese, and Martz [HWRM08] is a valuable recent addition to the reliability literature that treats, from a Bayesian point of view, many of the topics described in this paper. They describe the basic methodology for each area and illustrate the methods with a wide range of applications. Aven [Ave03], Singpurwalla, [Sin06], and Garrick [Gar08] use Bayesian frameworks in risk assessment applications.

There are many particular areas of application where the use of Bayesian methods is particularly compelling. As mentioned in Section 3, engineers often assume that some of the needed inputs to their reliability model are known, when in fact, there is always some degree of uncertainty. Use of a prior probability distribution to quantify knowledge would be better than incorrectly assuming that something is known without error.

The example in Chapter 14 of Meeker and Escobar [ME98] shows how bringing in a little prior information about the Weibull shape parameter (based on previous experience with the same failure mechanism in similar components) provided a much improved and more useful inference on reliability.

For another example, when doing temperature-accelerated life testing (like that described in Section 4) to accelerate a chemical reaction leading to failure, there are two

vastly different approaches that have been used. The use of a *known* activation energy in electronic component reliability modeling is common, especially in temperature accelerated testing of microelectronic components. This is, in effect, specifying the slope of the relationship between log lifetime and reciprocal absolute temperature (the Arrhenius model from physical chemistry). For example, MIL-STD-883 [MIL85] specifies testing at only one accelerated level of temperature and requires one to input a value of the activation energy (slope) in order to make reliability predictions, as described on page 282 of Nelson [Nel04]. Of course, the activation energy is not known exactly and assuming that it is known gives a false sense of having little statistical uncertainty (e.g., compare the analyses depicted in Figures 19.13 and 19.15 of Meeker and Escobar [ME98]). At the other extreme, the standard non-Bayesian approach would use the available data to estimate the activation energy as the slope of a linear relationship between life and reciprocal absolute temperature. Actually, there is, in most applications, useful but imprecise information about activation energy. An appropriate compromise analysis would use a prior distribution to describe the available information about activation energy. Section 22.2 of Meeker and Escobar [ME98] provides a simple example.

In all of these examples, either of the two extremes (assuming that the parameter is known versus assuming that nothing is known about the parameter) gives wrong answers, especially with respect to statistical uncertainty. Bayesian methods, however, provide an appropriate compromise.

Efron [Efr86] asked: “Why isn’t everyone a Bayesian?” If we rephrase the question to “Why isn’t everyone using Bayesian statistical methods in reliability applications?” I would assert that there are two primarily reasons:

- Concerns about the use of prior information.
- Lack of user-friendly software that would make the use of Bayesian methods easy.

Many of us are concerned with the use of Bayesian methods when there is the possibility of having wishful thinking or politically-driven opinions masquerading as legitimate prior information. Efron [Efr86] concluded that “The high ground of scientific objectivity has been seized by the frequentists.” Specification of a prior distribution would not be a primary concern if those at risk in a decision-making problem always have the opportunity to specify the prior (or to choose a diffuse prior, indicating lack of information to be used in the analysis). When, however, there are multiple parties at risk and lack of agreement on the prior to use, the objectivity that we seek in using data and statistical methods becomes elusive. *Objective Bayesian* methods (e.g., Berger [Ber06]) that have been developed over the past couple of decades may provide a solution to the technical side of this problem. In effect, the objective Bayesian methods specify a diffuse prior that will result in inferential results that are similar to what one would obtain from a non-Bayesian method. The ideas of objective Bayesian methods can be adapted to allow specification of prior information only for those quantities where there is legitimate prior information (e.g., prior information for which all of the risk-takers have a consensus).

The use of Bayesian methods in reliability will quickly become standard practice at the point in time when there is a software package that has

- Bayesian methods for a wide range of important reliability methods.
- Default diffuse prior distributions, but the ability to easily specify legitimate prior information in certain dimensions, when it is available.

- A carefully-designed graphical user interface that makes the package easy-to-use and hard to abuse.

12 Concluding remarks

Probability modeling and statistical methods are important tools in the reliability discipline. The new focus, for many manufacturers, in probabilistic design, assures that this importance will increase with time. Continuing changes in technology will assure that there will be important opportunities for statisticians to continue to make important research contributions to the discipline.

Acknowledgments: I would like to thank Yili Hong, Katherine Meeker, Mikhail Nikouline, and Ying Shi for helpful comments on an earlier version of this paper.

References

- [ABMR83] Abernethy, R. B., Breneman, J. E., Medlin, C. H., and Reinman, G. L. (1983). *Weibull Analysis Handbook*, Air Force Wright Aeronautical Laboratories Technical Report AFWALTR832079. Available from the National Technical Information Service, Washington, DC.
- [Ave03] Aven, T. (2003), *Foundations of Risk Analysis*, New York: John Wiley & Sons.
- [Ber06] Berger, J. (2006), The Case for Objective Bayesian Analysis, *Bayesian Analysis*, 1, 385-402.
- [Buc64] Buckland, W.R. (1964), *Statistical Assessment of the Life Characteristic*, London: Griffin.
- [Con93] Condra, L. W. (1993), *Reliability Improvement with Design of Experiments*, New York: Marcel Dekker.
- [CL07] Cook R. J. and Lawless J. F. (2007). *The Statistical Analysis of Recurrent Events*, New York: Springer.
- [CO84] Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman & Hall.
- [CKSS91] Crowder, M. J., Kimber, A. C., Smith, R. L., and Sweeting, T. J. (1991), *Statistical Analysis of Reliability Data*, New York: Chapman & Hall.
- [Cro01] Crowder, M. J. (2001), *Classical Competing Risks*, New York: Chapman & Hall.
- [DM78] David, H. A., and Moeschberger, M. L. (1978), *The Theory of Competing Risks*, London: Griffin.
- [Efr86] Efron, B. (1986), Why isn't everyone a Bayesian? (with discussion), *The American Statistician*, 40, 1-11.
- [ET93] Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.
- [EM99] Escobar, L.A. and Meeker, W.Q. (1999). Statistical prediction based on censored life data, *Technometrics*, 41, 113-124.
- [EMKK03] Escobar, L.A., Meeker, W.Q., Kugler, D.L. and Kramer, L.L., (2003), Accelerated Destructive Degradation Tests: Data, Models, and Analysis. In : *Mathematical and Statistical Methods in Reliability*, B. H. Lindqvist and K. A. Doksum, Editors, World Scientific Publishing Company.
- [FRY98] Flehinger, B., Reiser, B., and Yashchin, E. (1998) Survival with competing risks and masked causes of failure, *Biometrika*, 85, 151-164.

- [Gar08] Garrick, B. J. (2008), *Quantifying and Controlling Catastrophic Risks*, Amsterdam: Elsevier.
- [GK69] Gertsbakh, I. B., and Kordonsky, K. B. (1969), *Models of Failure, English translation from the Russian version*, New York: Springer-Verlag.
- [HM00] Haldar, A. and Mahadevan, S. (2000), *Reliability Assessment Using Stochastic Finite Element Analysis*, New York: John Wiley & Sons.
- [HWRM08] Hamada, M. S., Wilson, A. G., Reese, C. S., and Martz, H. F. (2008). *Bayesian Reliability*, New York: Springer.
- [ICS01] Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001), *Bayesian Survival Analysis*, New York: Springer.
- [JM00] Jeng, S. L. and Meeker W. Q. (2000), Comparisons of Weibull distribution approximate confidence intervals procedures for Type I censored data. *Technometrics*, 42, 135-148.
- [Law82] Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley & Sons.
- [Law03] Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, Second Edition, New York: John Wiley & Sons.
- [LC04] Lawless, J. F. and Crowder, M. (2004). Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, 10, 213-227.
- [LF05] Lawless, J. F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions, *Biometrika*, 92 , 529-542.
- [LM93] Lu, C. J., and Meeker, W. Q. (1993), Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35, 161-174.
- [MSS74] Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974), *Methods for Statistical Analysis of Reliability and Life Data*, New York: John Wiley & Sons.
- [MW82] Martz, H. F., and Waller, R. A. (1982), *Bayesian Reliability Analysis*, New York: John Wiley & Sons.
- [Mee87] Meeker, W. Q. (1987) Limited failure population life tests: application to integrated circuit reliability. *Technometrics*, 29, 151-165.
- [MD81] Meeker, W. Q. and Duke, S. D. (1981), CENSOR-A user-oriented computer program for life data analysis. *The American Statistician*, 35, 112.
- [ME98] Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, John Wiley and Sons, New York.
- [MEH09] Meeker, W. Q., Escobar, L. A., and Hong, Y. (2009), Using accelerated life tests results to predict field reliability. *Technometrics*, 51, xxx-xxx (in press).
- [MEL98] Meeker, W. Q., Escobar, L. A., and Lu, C. J. (1998), Accelerated degradation tests: modeling and analysis, *Technometrics*, 40, 89-99.
- [MIL91] MIL-HDBK-217F (1991), *Reliability Prediction for Electronic Equipment*. Available from Naval Publications and Forms Center, 5801 Tabor Ave, Philadelphia, PA 19120.
- [MIL85] MIL-STD-883 (1985), *Test Methods and Procedures for Microelectronics*. Available from Naval Publications and Forms Center, 5801 Tabor Ave, Philadelphia, PA 19120.
- [Mur93] Murray, W. P. (1993), Archival life expectancy of 3M magneto-optic media, *Journal of the Magnetics Society of Japan*, 17, Supplement S1, 309-314.
- [Nel82] Nelson, W. (1982). *Applied Life Data Analysis*, New York: John Wiley & Sons.
- [Nel90] Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, New York: John Wiley & Sons
- [Nel03] Nelson, W. (2003). *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*, ASASIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, PA.

- [Nel04] Nelson, W. (2004). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, New York: John Wiley & Sons (updated paperback version of the original 1990 book).
- [NH72] Nelson, W., and Hendrickson, R. (1972), *1972 User manual for STATPAC—a general purpose program for data analysis and for fitting statistical models to data*. General Electric CR&D Technical Report 72GEN009.
- [RH04] Rausand, M., and Hoyland, A. (2004) *System Reliability Theory: Models and Statistics Methods*, (Second Edition) New York: John Wiley & Sons.
- [Sin06] Singpurwalla, N.D., (2006) *Reliability and Risk: A Bayesian Perspective*, New York: John Wiley & Sons.
- [Str80] Strauss, S. (1980), STATPAC: A general purpose program for data analysis and for fitting statistical models to data, *The American Statistician*, 34, 59-60.
- [VM90] Vander Weil, S., and Meeker, W. Q. (1990), Accuracy of approximate confidence bounds using censored Weibull regression data from accelerated life tests. *IEEE Transactions on Reliability*, 39, 346-351.
- [WM02] Wu, H., and Meeker, W. Q. (2002), Early detection of reliability problems using information from warranty data bases. *Technometrics*, 44, 120-133.