Statistical Methods in Reliability

J. F. Lawless

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Some of the advances made during the past 25 years in the statistical treatment of reliability problems are reviewed. The impact of statistical methods on reliability is discussed, and some areas where work is needed are suggested.

KEY WORDS: Lifetime data; Failure data; Estimation.

1. INTRODUCTION

This paper attempts to review some of the advances made during the past 25 years in the statistical treatment of reliability problems, to assess their impact, and to suggest some areas where work is needed. The views taken are of course personal ones, and though I have benefited greatly from the comments and insights of numerous people, responsibility for the opinions expressed is mine alone.

I use the term reliability here to refer to that field that deals with the study of the proper functioning of equipment and systems, and the design and production of such equipment. I deal only with statistical methods in reliability; other mathematical and probabilistic aspects of reliability (see e.g., Barlow and Proschan 1965; Shooman 1968) are considered only in as much as they relate to the statistical methodology being discussed. This narrows the focus of the paper considerably, since there is an enormous literature on mathematical models and aspects of reliability.

I decided to focus on development during the past 25 years, partly because this spans the life of this journal, but also because time and space limitations make it difficult to cover a longer period in any detail. I review developments prior to 1959 in Section 2 to provide a brief historical perspective on statistical methods in reliability. Section 3 considers some methodological advances of the past 25 years, with a brief assessment of the work. Section 4 considers the impact of statistical methods on the field of reliability. Section 5 proposes some areas for future development, and Section 6 concludes the paper.

2. HISTORICAL PERSPECTIVE

The field of reliability is of recent origin. With the widespread manufacture and use of increasingly sophisticated mechanical, electrical, and electronic equipment during this century, questions of reliability became of interest. Before 1940, most related work concerned either quality control or machine maintenance problems, and reliability was not identified as a specific field. However, with the Second World War and the advent of more and more sophisticated equipment in the military and elsewhere, reliability came into its own. Modern reliability, and the identification of reliability as a specific field, came into being in the late 1940's and early 1950's. In the early 1950's, reliability engineering blossomed, and several groups began formal studies of reliability problems; these have had a long-lasting effect on the statistical treatment of the area. The best-known example is perhaps the Advisory Group on the Reliability of Electronic Equipment (AGREE), formed by the U. S. Department of Defense in 1952, which subsequently developed widely used specification standards for the reliability of electronic equipment.

Much of the statistical work in reliability has focused on the lifetime of pieces of equipment or systems, or on their failure or nonfailure over a period of time. Relatively little seems to have been done in this area until after about 1935 (but see Kurtz 1930), and
most of the early work tended to use actuarial methods of handling lifetime data. In the late 1930's and the 1940's, parametric families of distributions that could be used as lifetime models began to appear in engineering contexts; some of these, such as the extreme value distributions (Gumbel 1935) and the Weibull distribution (Weibull 1939), appeared first in connection with the strength or fatigue life of materials. Statistical analyses of lifetime data by nonparametric (especially actuarial) methods also continued to appear (e.g., Altman and Goor 1946). Then, in the early 1950's, there coincided with the general increase of attention to reliability problems by engineers an increased interest in the statistical treatment of reliability data. The use of the exponential distribution dominated much of the early work. Davis (1952) wrote an influential early paper on the use of the exponential as a life distribution, and Epstein and Sobel (e.g., 1953,1954) and others (e.g., Bartholomew 1957) provided statistical methods for it, for both censored and uncensored data. This work was soon adopted by reliability engineers: the 1957 AGREE report, which led to the well-known MIL-S-1117-781 reliability test standard, considered essentially everything in terms of the exponential distribution.

In the mid-to-late 1940's, other distributions began to appear frequently in life distribution work. The Weibull distribution, especially, became popular: Weibull (1951), Kao (1958), and Lieblein and Zelen (1956) were all influential in popularizing the distribution and in starting to provide statistical methods for it.

The situation around the end of 1958 was thus roughly as follows: Reliability had taken off as a major area in engineering. The exponential was well established as the distribution on which much statistical analysis and modeling of lifetimes, and hence of reliability, was based. The exponential was particularly popular with reliability engineers, though some other distributions, particularly the Weibull, were starting to gain use. Only for the exponential distribution were statistical methods reasonably well developed; even here, most of the new work had focused on one- or two-sample problems with censored data, and other techniques such as regression analysis had not been very thoroughly considered.

Two other developments were also taking place at this time. First, work was starting on the estimation of system reliability: Buehler (1957) and Steck (1957), for example, had estimated reliability for a series system in the binary context (i.e., each component either fails or does not fail over some period of time), based on test data on component reliabilities. Second, mathematicians and reliability engineers were starting to consider more sophisticated models for systems and were introducing concepts like availability and maintainability (e.g., see Shooman 1968). The reliability growth of systems under development was also studied and modeled. These ideas provided challenges to the statistician although, as we shall see, they have not led to as many statistical advances as one might have hoped for.

3. SOME STATISTICAL DEVELOPMENTS DURING 1959–1983

The past 25 years have seen numerous advances in statistical theory and methods relevant to reliability. I review what I see as some of the main developments under six headings, which are by no means mutually exclusive or exhaustive. I occasionally comment on the application and impact of the methodology, and on where more work is needed. In Sections 4 and 5, I discuss these issues in somewhat broader terms.

3.1 Parametric Inference for Univariate Life Distributions

One of the major areas of activity in the 1960's and 1970's was in the provision of estimation and hypothesis test procedures for various parametric life distribution families, for both censored and uncensored data. Books by Bain (1978), Lawless (1982), Mann, Schäfer, and Singpurwalla (1974), and Nelson (1982) review much of the work in this area; I deal only with a few major aspects.

Before 1959 a considerable amount of work had been done on inference procedures for the exponential distribution with censored and uncensored data (e.g., Epstein and Sobel 1953,1954; Bartholomew 1957) and for the normal and lognormal distributions. However, relatively little had been done with methodology for censored data in general and for distributions such as the Weibull, gamma, and lognormal (with censored data). Some of the impetus for the great amount of activity in this area in the 1960's and 1970's came from the realization that whereas the exponential distribution was easily handled statistically, it gave nonrobust methods (e.g., Zelen and Dannemiller 1961) and was not an appropriate model in many situations. At the same time, other life distributions such as the Weibull were becoming increasingly popular, and methodology was needed for them.

The Weibull emerged in the 1960's and 1970's as perhaps the most widely used life distribution. Research on it was particularly heavy during this period and is worth examining a bit more closely for several reasons, one being that many of the statistical methods developed for this model are now routinely used in life testing and reliability work.

The Weibull distribution has probability density function (pdf) of the form

\[ g(t; \alpha, \beta) = \frac{\beta}{\alpha} \left( \frac{t}{\alpha} \right)^{\beta - 1} \exp \left[ -\left( \frac{t}{\alpha} \right)^\beta \right], \quad t > 0, \]  

(3.1)
where \( x > 0 \) and \( \beta > 0 \) are scale and shape parameters, respectively. When lifetime \( T \) has pdf (3.1), log lifetime \( X = \log T \) has an extreme value distribution with pdf

\[
f(x; u, b) = \frac{1}{b} \exp \left\{ \frac{x - u}{b} - \exp \left[ \frac{(x - u)}{b} \right] \right\},
-\infty < x < \infty
\]

where \( u = \log a \) and \( b = \beta^{-1} \) are location and scale parameters, respectively. In the 1950’s, Kao (e.g., 1958), Lieblein and Zelen (1956), and others had considered point estimation of \( x \) and \( \beta \) (or, equivalently, \( u \) and \( b \)), but nothing much had been done relating to interval estimation or hypothesis testing. In the late 1960’s and early 1970’s however, major advances in this area were made by Johns and Lieberman (1966), Mann (e.g., 1968) and her colleagues (e.g., Mann and Fertig 1973), Bain and his colleagues (e.g., Thoman, Bain, and Antle 1969; Billman, Antle, and Bain 1972), and others. All of this work rested on the fact that for estimators \( \hat{u}, \hat{b} \) of \( u \) and \( b \) in (3.2), which are equivariant (see Lawless 1982, p. 533), quantities such as \((\hat{u} - u)/b, \hat{b}/b\) and \((\hat{u} - u)/\hat{b}\) are pivotal (parameter-free), and their distributions can be used to obtain (conditional or unconditional) confidence intervals or hypothesis tests for \( u, b \), and certain functions of \( u \) and \( b \). One of the interesting things about this is that the basic statistical theory had been known since the time of Fisher’s early work (e.g., Fisher 1934) but that implementation of both conditional and unconditional procedures for the extreme value distribution (3.2) had to await the availability of modern computing equipment. Another interesting point is that use of these techniques on the extreme value and other location-scale parameter distributions such as the logistic and log generalized gamma (e.g., Lawless 1980) has highlighted some of the aspects of so-called conditional versus unconditional inference procedures. In the context of the extreme value distribution, there is essentially no practical difference in the two approaches, though the outlooks of the approaches differ. Lawless (1978) reviews this.

As a word of criticism on the way in which statistical theory and methods often evolve, I feel that too much has been published on statistical inference for the Weibull distribution, as well as for the exponential. (I did a quick count of the papers in Technometrics from 1967-1980 that dealt in some way with statistical methods for reliability. I identified 179 such papers, of which 80 dealt with estimation or testing for the ordinary exponential or Weibull distributions.) Some of the effort spent on the Weibull and exponential distributions might have been more usefully directed to investigations of other models.

Concerning inference procedures for univariate life distributions, it should also be said that useful work on other models, such as the gamma (e.g., Engelhardt and Bain 1978), the generalized gamma (e.g., Farewell and Prentice 1977; Lawless 1982, Sec. 5.3), the lognormal (e.g., Nelson and Schmee 1979), and the inverse Gaussian (e.g., Chhikara and Folks 1977) has been done in recent years, as has work on goodness-of-fit tests for various models. In addition, our understanding of general large-sample maximum likelihood methods for use with censored data has improved considerably. An early paper by Halperin (1952) addressed large-sample properties of maximum likelihood for the case in which the data are Type II censored (i.e., when one observes only the \( r \) smallest observations in a random sample of size \( n \)). Later work by others such as Blight (1970), Cox (1975), Aalen (1978), and Kalbfleisch and Prentice (1980, Ch. 5), some of it motivated by biomedical lifetime problems, has established that standard maximum likelihood large-sample methods can be used with many types of censored data. In particular, if one has a sample from a life distribution with pdf \( f(t; \theta) \) and corresponding survivor, or reliability, function \( S(t; \theta) = \int_{t}^{\infty} f(x; \theta) \, dx \), consisting of observed lifetimes \( t_1, \ldots, t_k \) and right-censoring times \( t_{k+1}, \ldots, t_{n+1} \), the observed likelihood function for \( \theta \) can under fairly broad conditions be taken to be

\[
L(\theta) = \prod_{i=1}^{k} f(t_i; \theta) \prod_{i=k+1}^{n} S(t_i; \theta).
\]

In addition, \( \hat{\theta} \), obtained by maximizing \( L(\theta) \), and the likelihood itself provide inference procedures for \( \theta \) through standard large-sample use of likelihood ratio statistics, or approximate normality of \( \sqrt{n}(\hat{\theta} - \theta) \).

It should also be remarked that although large-sample procedures can be routinely applied with most parametric life distributions, there is still a need for specialized investigations of methods for specific models. A main reason for this is that in many reliability applications the number of observed lifetimes is not sufficiently large for one to be sure that the asymptotic distributional approximations used are truly accurate. Work remains to be done in this area for all of the popular lifetime distributions. More investigation of other special problems, such as the analysis of grouped (interval) data from various distributions, is also needed.

### 3.2 Nonparametric and Graphical Procedures With Censored Data

Another area where advances have been made in lifetime data analysis is in the extension of nonparametric procedures to handle censored data. I mention advances made in two directions. First, a major accomplishment, though it falls just outside of the period 1959–1983, was the development of the
product-limit estimate of the survivor function $S(t) = \Pr (T \geq t)$ for a lifetime random variable $T$, by Kaplan and Meier (1958). Closely related to this is the nonparametric estimate of the cumulative hazard function $H(t) = -\log S(t)$ given by Nelson (1972a) and Altschuler (1970). The product-limit estimate is defined as follows: suppose that $t_1 < t_2 < \cdots < t_k$ are the observed lifetimes (for simplicity, assumed here to be distinct) in a sample of $k$ lifetimes and $n - k$ censoring times. Let $n(t)$ denote the number of individuals known to be still alive (i.e., alive and uncensored) just prior to time $t$. Then the product-limit estimate of $S(t)$ is

$$
\hat{S}(t) = \prod_{i: t_i < t} \left( \frac{n(t_i) - 1}{n(t_i)} \right),
$$

(3.3)

with the proviso that if there happens to be a censoring time $t^*$ larger than $t_k$, then $\hat{S}(t)$ is undefined past $t^*$. The Nelson (1972a) estimate of $H(t) = -\log S(t)$ is

$$
\hat{H}(t) = \sum_{t_i < t} \frac{1}{n(t_i)},
$$

(3.4)

which yields another estimate $\hat{S}(t) = \exp [-\hat{H}(t)]$ of $S(t)$. Breslow and Crowley (1974) and others have examined asymptotic properties of (3.3) and (3.4).

Besides giving nonparametric estimates of a life distribution's survivor function from censored data, (3.3) and (3.4) are extremely useful for providing plots from which one can carry out model assessment or rough parameter estimation for parametric families of models (e.g., Nelson 1982, Chs. 3,4; Lawless 1982, Ch. 2).

In another direction, nonparametric tests for the equality of two or more distributions were extended to handle censoring. The best-known examples are perhaps the extensions of the Wilcoxon test (e.g., Efron 1967; Prentice 1978) and the exponential ordered scores test (e.g., Peto and Peto 1972; Prentice 1978). There is much recent additional work in this area—for example, concerning the power of various tests in different situations.

Graphical methods are widely used by statisticians and engineers in reliability work, and the estimates (3.3) and (3.4) and related probability and hazard plots are particularly useful (e.g., see Nelson 1982, Chs. 3,4). On the other hand, the nonparametric tests for equality of distributions were mostly developed in response to lifetime problems in biomedicine, and have not been used much in reliability. There has been a marked preference in the reliability area, especially among engineers, for parametric distribution-based methods. This is perhaps a reflection on the types of lifetime data problems often considered in reliability, but it is also true that there is not enough awareness of the pitfalls associated with parametric methods (e.g., see Hahn and Meeker 1982), and somewhat more frequent use of nonparametric methods is likely warranted. Nonparametric methods may also be increasingly useful in studies of field (as opposed to laboratory) data; see point 1 in Section 5.2.

### 3.3 Regression Analysis of Lifetime Data

In the last 25 years it has become routine to handle regression problems in which lifetime, or some transform thereof, is the response variable. Maximum likelihood large-sample theory or linear estimation (including least squares) methods are typically used. Such methods have been developed and applied in numerous papers; Nelson and Hahn (1972), Nelson (1972b), Lawless and Singhal (1980), and Zelen (1959) provide examples in the life-testing area. Applications to accelerated life testing are especially common.

The regression models used most with lifetime data fall into two categories. First, most of the common parametric models are of the form

$$
Y = \mu(x; \beta) + \sigma Z, \quad -\infty < Z < \infty
$$

(3.5)

where $Y = \log T$ represents log lifetime, $\mu(x)$ is a function of a vector of regressor variables $x = (x_1, \ldots, x_p)$, $\sigma$ is a positive scale parameter, and $Z$ has a known distribution. The most widely used models are the extreme value, where $Z$ has pdf $\exp (z - e^z)$, and the normal, where $Z$ has a standard normal distribution. These models correspond to lifetime $T$ given $x$ having Weibull and lognormal distributions, respectively. The form also includes the exponential distribution as a special case. Models of the form (3.5), though not always written this way, are used in connection with lifetimes of many types of items (e.g., see Lawless 1982, Ch. 6; Mann, Schafer, and Singpurwalla 1974, Ch. 9).

The second category of models, much used in the regression analysis of biomedical lifetime data, is the so-called proportional hazards family. Here, the effect of regressor variables is assumed to be multiplicative on the hazard function for $T$, which is given by $h(t) = f(t)/S(t)$. The hazard function of $T$ given $x$ is assumed to be of the form

$$
h(t | x) = h_0(t; \theta)g(x; \beta),
$$

(3.6)

where $\theta$ and $\beta$ are parameters. Here, $h_0(t; \theta)$ can be thought of as the baseline hazard function for an item with $g(x; \beta) = 1$. Another way to write (3.6) is in terms of the survivor function of $T$ given $x$ as

$$
S(t | x) = [S_0(t; \theta)]^{g(x; \beta)},
$$

(3.7)

where $S_0(t; \theta)$ is now the survivor function of an item for which $g(x; \beta) = 1$.

Fully parametric models of the form (3.6) are sometimes used by selecting a specific family of baseline hazard functions $h_0(t; \theta)$. The model (3.5) in which $Z$ has an extreme value distribution in fact turns out to be of this type and corresponds to $T$ given $x$ always
having a Weibull distribution. A more common approach to statistical analysis with these models, however, is a semiparametric one due to Cox (1972) in which a parametric form is assumed for \( g(x; \beta) \) but \( h_0(t; \theta) \) is left completely arbitrary. Cox (1972) has perhaps influenced the analysis of lifetime data in the biomedical area more than any other paper in the last 25 years; the book by Kalbfleisch and Prentice (1980) is the key source for this area of methodology. As with nonparametric techniques in the nonregression case, however, these methods have to date scarcely been used in the reliability area. Again, it is my feeling that such methods may see increased usage if more emphasis is given in reliability to the analysis of field data. In addition, the recent availability of computer programs for these methods, for example in SAS and BMDP, may increase their popularity in the reliability area.

Regression methods based on binary response models also deserve brief mention, since these are frequently used when one considers only the failure or nonfailure of equipment over a specified time period. Such models have been used for a long time in connection with dose-response problems in biology, but in the past 20 or 30 years they have also been widely used in reliability and elsewhere. One can of course adapt the models (3.5) or (3.6) to this situation; most of the models used are adaptations of (3.5). In particular, (3.5) gives, for a specified time \( t_0 \), the probability of failure by time \( t_0 \) as

\[
\Pr(T \leq t_0) = \text{Pr}
\left( \frac{Y - \mu(x; \beta)}{\sigma} \leq \frac{\log t_0 - \mu(x; \beta)}{\sigma} \right) = F\left(\frac{\log t_0 - \mu(x; \beta)}{\sigma}\right),
\]

where \( F \) is the cumulative distribution function (cdf) of \( Z \). Frequently \( \mu(x; \beta) \) is taken to be a linear function; in the case of a single regressor variable this would be \( \mu(x; \beta) = \beta_0 + \beta_1 x \), and (3.8) would become

\[
\Pr(\text{failure by } t_0) = F(\alpha + \beta x),
\]

where \( \alpha = (\log t_0 - \beta_0)/\sigma \) and \( \beta = -\beta_1/\sigma \). The most frequently used models are those for which \( F \) is the cdf of a standard logistic, standard normal, or standard extreme value distribution. Cox (1970) discusses models of this sort in general, and Nelson (1982, Ch. 10) discusses them in a reliability context.

**Remark.** Sections 3.1 to 3.3 deal with statistical techniques for univariate life distributions, including the regression case. These methods are fundamental to reliability work, and they are widely and for the most part appropriately used. Many reliability problems, however, require a different framework. Statistical models and methods have been developed for many situations, but much remains to be done, both in the development and in the application of suitable methodology. Sections 3.4 to 3.6 briefly examine three problem areas.

### 3.4 Multiple Failure Mode Problems

Many problems dealing with equipment involve more than one type of failure. Often, there is associated with each item a pair of variables \((T, C)\), where \( T \) is lifetime and \( C \) denotes a failure mode, or a cause of failure. Let us assume that \( C \) takes on values in \( \{1, 2, \ldots, k\} \). The main approach to this problem has been through the classical theory of competing risks (e.g., see David and Moeschberger 1978; Nelson 1982, Ch. 5). In this formulation, one supposes that there is associated with each of \( k \) possible failure causes \((i = 1, \ldots, k)\) a random variable \( T_i \) that represents time to failure from that cause. The time to failure for the item is then given by \( T = \min(T_1, \ldots, T_k) \), and \( C = \{i: T_i = \min(T_1, \ldots, T_k)\} \) is the observed failure cause. Most competing risk analysis assumes that \( T_1, \ldots, T_k \) are independent. In this case the survivor function for \( T \) is \( S(t) = S_1(t) \cdots S_k(t) \), where \( S_i(t) \) is the survivor function for \( T_i \). Either parametric or nonparametric statistical inference about \( S_i(t) \), \( S_k(t) \) is easily developed, since observing one cause of failure, \( j \), on a given item merely means that \( T_j \) is observed and the remaining \( T_i \)'s \((i \neq j)\) are censored (i.e. we know only that \( T_i > T_j \) for each). David and Moeschberger (1978) and Nelson (1982, Ch. 5 and Section 5.5 of Ch. 8), for example, give details of this analysis.

For systems consisting of \( k \) separate components in series, the classical framework is generally appropriate, and the assumption of independence is also often warranted. (Then, \( T_1, \ldots, T_k \) represent the lifetimes of the \( k \) components.) However, the classical approach is frequently forced on many other situations where it is inappropriate. A more useful approach for many situations, developed over the past few years, is mentioned later in this section.

A main problem with the classical approach (outside of the series system problems mentioned previously) is that the concept of \( T_i \), time to failure from cause \( i \), is often without any physical basis. In addition, even when the \( T_i \)'s are physically meaningful, they are often not independent. A frequently suggested alternative is to use a parametric multivariate distribution for \((T_1, \ldots, T_k)\) that allows dependence (e.g., David and Moeschberger 1978, Ch. 4). However, the work of Cox (1959), Tsiatis (1975), Peterson (1976), and others shows that it is not possible to test the validity of any such model if one observes only data of the form \((T, C)\). In fact, for any model assuming dependent \( T_i \)'s there is always a model with independent \( T_i \)'s giving the same distribution for \((T, C)\) (e.g.,
Lawless 1982, p. 483). Another difficulty with the classical competing risks formulation is that it tempts people to treat the effect of removing one or more causes of failure by merely deleting the corresponding \( S_i(t) \) 's in \( S(t) = \prod_{i=1}^{k} S_i(t) \). This is almost never appropriate.

An alternative formulation of multiple failure mode problems is given by Altschuler (1970), Prentice et al. (1978), and others. Briefly, this models the joint distribution of \((T, C)\) directly, in terms of cause-specific hazard functions

\[
h_j(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t, C = j | T \geq t)}{\Delta t}
\]

or subsurvivor functions \( S_j(t) = Pr(T \geq t, C = j) \), \( j = 1, \ldots, k \). With this approach both parametric and nonparametric inference procedures are easily developed (e.g., see Lawless 1982, pp. 484–491). In cases in which the classical competing-risks framework is physically meaningful and \( T_1, \ldots, T_k \) are independent, the formulation just given is equivalent to the competing-risks framework and so yields the same statistical procedures. In other situations the present formulation seems much more satisfactory.

### 3.5 Methods for Repeated Failure Analysis

An important reliability problem is the treatment of repeated failures on the same piece of equipment or system; for an oft-discussed example see Proschan (1963), where data on successive failures of air conditioning equipment in airplanes are given. In the past 25 years there has been in statistics generally a wider use of stochastic point process models and related statistical methods; Cox and Lewis (1966) and Lewis (1972) are two key references in the area. However, though there has been a lot of activity in applying point process and related ideas in modeling reliability problems, there has been much less work on model fitting and statistical analysis. Recently, there has been more activity in this direction (e.g., Ascher and Feingold 1978, 1983; Bain and Engelhardt 1980; Barlow 1978; Crow 1982; Lee 1980). Much, however, remains to be done in developing and applying statistical methods with stochastic process models in reliability and in making existing methods better known. Ascher and Feingold (1983) discuss this area in detail.

Even in quite simple situations there is room for work. For example, consider repeated failure of a piece of equipment over time, with repair time for simplicity being ignored. One way to model this is in terms of a point process (e.g., Cox and Miller 1965, Ch. 9) with intensity function

\[
\lambda(t; H(t), x)
\]

where \( H(t) \) represents the past history of failures for the equipment, and \( x \) represents concomitant variables that might be related to failure. (These covariables might also depend on time, but for simplicity here I'll assume they are fixed.) The key problem is to develop models and statistical methods for (3.10). Two models that are widely used in reliability are the time-dependent Poisson process for a homogeneous population, where \( \lambda(t; H(t), x) = \lambda(t) \), and the renewal process, where \( \lambda(t; H(t), x) = \lambda(t - \nu(t)) \), with \( \nu(t) \) the time of the most recent failure. However, there needs to be more done in developing models and in discriminating among alternative models. Some recent work in this direction in a biomedical context is given by Prentice, Williams, and Peterson (1981). Existing methodology also needs to be better publicized; on this account, see Ascher and Feingold (1978, 1983).

Another key point is that models of the form (3.10) generally require a quite different point of view than do models for the life distribution (or time to a single failure) of a piece of equipment. Perhaps because of the large amount of work on life distribution methodology, and also because of the intimate connection between the homogeneous Poisson process and the exponential distribution, there has been a lot of confusion of the two sets of ideas in the reliability literature; Ascher and Feingold (1978, 1983) and Thompson (1981) discuss this confusion. Ascher and Feingold (1983) present an excellent detailed critique of reliability work with respect to neglect of point-process statistical methods for repairable systems.

### 3.6 Estimation of System Reliability

The determination of system reliability from reliability information on components of the system has been discussed a great deal (e.g., Barlow and Proschan 1975), but statistical methods have played a fairly minor role. A typical statistical problem is the following: suppose that a system has components \( C_1, \ldots, C_k \) connected in some configuration. Given data from independent tests on the various components, estimate the reliability of the system. Now, estimation of the failure time distributions of components is well taken care of (see Section 3.1), but the estimation of system reliability from this (say, in terms of a confidence interval for probability of successful operation at time \( t \)) is usually not easy.

A lot of effort has gone into developing estimation procedures for systems in which the components \( C_1, \ldots, C_k \) operate independently of one another. Mann, Schafer, and Singpurwalla (1974, Sec. 10.4) review work in this area but essentially, methods are available only for exponential failure time distributions, or binary (fail/no-fail) treatments of components and systems.

The assumption of component independence is of course critical to the work just mentioned, and this
limits its applicability. There are also obvious difficulties in trying to estimate the reliability of a system in actual operation, on the basis of laboratory test data. Consequently, besides the extension of existing approaches to wider ranges of life distributions, there is a need for models that capture more of the physical realities of system reliability problems. For this work to be useful, however, there also has to be a strong emphasis on the collection of reliability data suitable for model assessment.

4. COMMENTS ON THE IMPACT OF STATISTICAL METHODS IN RELIABILITY

The area in which statistics has had its greatest impact in reliability is in the analysis of laboratory and field data on lifetimes and failure. Life test data for items such as ball bearings, electrical insulation, and semiconductor devices are commonly analyzed under parametric models such as the Weibull and lognormal distributions. Statistical life test and reliability demonstration procedures have also been written down as mandatory standards (e.g., MIL-STD-781C, 1979), to be followed in certain situations. The development of estimation and hypothesis testing methods for numerous parametric lifetime distributions in the 1950's, 1960's, and 1970's, referred to in Section 3, was instrumental in making this possible.

Not all that has happened with lifetime data analysis in reliability has been good. For one thing, the statistical approach to lifetime or failure problems has often been very stereotyped. The exponential distribution and methods based on it have, for example, been much more widely used than they should be. This has also carried over to the treatment of more complicated system reliability problems, where the exponential distribution and related homogeneous Poisson process are greatly overused. One factor contributing to this is that many engineers resist any method that they see as complicated, preferring to work with a few standard procedures. Statisticians have not been particularly successful at getting them to take a more broadly based approach to data analysis. At the same time, statisticians have perhaps concentrated too much in the literature on statistical niceties for certain distributions, and too little on innovative methods of life data analysis.

Outside of the area of life data analysis, the impact of statistics in reliability has in my opinion been rather limited. One reason may be simply that not enough statisticians are actively involved with real reliability problems. Although there is a great deal of mathematical and probabilistic modeling done in reliability (e.g., see books such as Barlow and Proschan 1965, 1975; Shooman 1968), there is not enough statistical work relating it to real-world data. In a book review, Easterling (1977) remarks on this when he says,

All is not well among reliability mathematicians and engineers. The engineers fault the mathematicians for developing models which cannot be implemented and the mathematicians fault the engineers for unwillingness to attempt to understand and apply new mathematical results. Both can find support for their positions in this volume. Largely missing from the mix... is the role of the statistician, the data analyzer, the model builder.

This is not to say that statistics has had no impact beyond the area of life data analysis. There are many interesting and sensible applications of statistics to be found in the statistics literature, in subject-matter journals and conference proceedings, and in internal reports. For example, sound statistical methods for dealing with repeated failure data (see Sec. 3.5), for assessing reliability growth, and for analyzing field data are all seen to be used, albeit not as frequently and not always as wisely as one would like. On the other hand, a quick look at an area such as safety and risk assessment (e.g., see the U.S. Nuclear Regulatory Commission 1975 study on nuclear reactor safety) shows not only that there are many challenges for the statistician, but that even sound basic statistics has not permeated much of the reliability/safety area.

Reliability is a large field that deals with problems in many branches of science and engineering, and I cannot address the impact of statistics in any specific area here. Furthermore, discussion of the wise use of statistical methods in a specific area of application always poses several delicate questions. However, I think it is fair to say that there is a need for statisticians to become more involved in the collection and analysis of reliability data and in the modeling aspects of reliability. In many situations, just the sensible application of basic methods would have a major impact, but there is also room for the development of new techniques. The next section looks at a few areas where work could be done.

5. SOME AREAS WHERE WORK IS NEEDED

I have mentioned in Section 3 a few places where new work might be done. In this section I indicate some broad areas where work is needed, and I discuss a number of specific problems. In some cases the statistical tools necessary to tackle the problems already exist; in others, new ideas are called for.

5.1 Taking Statistical Methodology to Reliability People

Useful statistical methodology exists for many reliability problems (the topics in Section 3 are good examples) but is not widely enough known or applied. In addition, there are numerous problems that engineers are just not sure how to handle statistically.
The following areas deserve more attention:

1. More statisticians need to be involved in real reliability problems, and engineers need to be better educated statistically. This is affected to a considerable degree by the number of graduate statisticians that find their way into government, military, and industrial reliability work. University statisticians could, however, (beyond trying to graduate more statisticians) (a) try to become involved in practical reliability work, say through visits to organizations that have reliability problems, or through university engineering faculties; and (b) consider whether more could be done in teaching statistics to reliability engineers and scientists. This includes the provision of short courses to practicing engineers and scientists, as well as (service) teaching within the university.

Incidentally, I do not mean to imply that only statisticians are capable of doing good statistics; this is not true. However, as a statistician writing in a statistics journal, I am preoccupied with challenges facing statisticians in the field of reliability.

2. Statistical methodology gets more widely and rapidly used if it is implemented by good computer programs. There are currently several program packages that do a variety of analyses for lifetime data; Nelson (1982) references most of them. Some of the packages, such as STATPAC (Nelson, Morgan and Caporal 1978), SURVREG (Preston and Clarkson 1980) and CENSOR (Meeker and Duke 1979) are fairly versatile, but there is at present no single package that (a) is portable, (b) analyzes censored data for the common life distributions, including estimation and goodness-of-fit tests, (c) handles regression analysis based on the common life distributions, (d) includes standard nonparametric procedures such as life table methods, and Cox's (1972) proportional hazards regression analysis (see Section 3.3), and (e) provides good graphics capabilities. Further work on a major package would be valuable.

Besides program packages with a wide range of lifetime data analysis capabilities, there is a need for good programs to do quite specific tasks. If such programs are to serve nonexpert statisticians, they should be particularly straightforward to use; there is a reluctance on the part of many engineers to employ statistical software unless it is extremely easy to do so. Statisticians might also publish more tutorial papers and software descriptions in reliability subject-matter journals.

3. There are numerous problems of practical importance that statisticians do not bother with much because they are "messy," or do not provide any interesting "new" methodology. Examples that come to mind are inference procedures for parametric life distributions based on grouped data, and analysis of variance techniques for censored data from distributions such as the Weibull. Statisticians provide a valuable service if they can give tidy methods for handling them, and it is especially useful to publish such methodology in subject-matter journals. McCool (1979) provides an example of this kind of work.

5.2 Data Collection Problems

There are major data collection problems in reliability, both in the design of experimental studies and in the collection of field data. I shall comment briefly on each area.

1. There should be more care given to the collection (and analysis) of field data. It is widely recognized that laboratory life test results frequently say little about the reliability of equipment in actual (field) use. When good field data are collected, more can be done to relate equipment or system reliability to the environment in which it operates, as well as to inherent properties of the equipment. Attempts can also be made to validate some of the models suggested for field reliability. Field data are collected now in many areas, and there even exist some very large banks of field failure data (see Gertz 1974 for one example). Often, however, the data are seriously incomplete: for example, only failure (and not nonfailure) information may be recorded; data on failures may be incomplete; no or little information may be collected on the environment in which the equipment was operating; there may be bias in the reporting of failures (e.g., only some failures are reported); and so on. The collection of good field reliability data poses challenging and important problems yet is virtually undiscussed in the statistical literature. Amster, Brush, and Saperstein (1982) is a recent exception.

2. In connection with the design of laboratory studies, statisticians should be continually pointing out the limitations of certain types of experiments. (For example, one practice whose value I often question is the use of step-stress accelerated life tests, for example on electrical insulation. Whether these help much either for comparing insulations or in assessing field reliability is often doubtful.) Statisticians should also encourage engineers and scientists to consider experiments that might lead to more useful statistical models and that would provide information on difficult statistical questions.

5.3 Statistical Analysis and Modeling Problems

There are many areas where work related to statistical models and data analysis could profitably be done. A few are mentioned here.

1. In practice, failure often occurs in ways that do not mesh with the point of view taken in most statistical work on lifetime data. For example, Blanks (1973) says, "...the major cause of electronic compo-
nt failure is quality defects, not systematic mecha-
isms inherent in the intrinsic device." However, most
of the statistical life distribution work is geared to
modeling systematic failure mechanisms, resulting in
gradual wearout and failure. There should be more
emphasis on mixture models, particularly ones that
directly consider early failure mechanisms. Hahn and
Meeker (1982, Part II) discuss some models of this
kind; also see Miller (1960).

Frequently the applicability of homogeneous
wearout models is justified by subjecting equipment to a
burn-in, thus eliminating components with failure-
cauing defects (e.g., Barlow et al. 1968). It might be
noted that burn-in has not been much studied in the
literature either, though a recent book on this area,
aimed at engineers, has been written by Jensen and
Petersen (1982).

2. There are several questions about how to ap-
proach situations in which failures can be of different
types. Possible scenarios include (a) each item is sus-
sceptible to only one type of failure (the mixture sit-
uation); (b) all items are subject to all types of failure
(the usual competing risks situation); (c) different
items may be subject to some, but not necessarily to
all, types of failure. To complicate matters, the scenar-
ios may depend on the environment in which the
equipment operates. Questions arise as to the proper
way to model specific multiple-mode failure situa-
tions; the so-called competing risks framework was
discussed in Section 3.4. Questions concerning the
type of data necessary to assess models adequately are
of particular interest.

3. More attention needs to be given to looking at
the various sources of variability contributing to life-
time and failure data. For example, Little (1980) re-
ports situations in which the variability in log fatigue
life of certain types of material among laboratories
was about eight times as large as the variability within
laboratories. Another source of variability is batch-to-
batch variability in manufactured equipment; Hahn
and Meeker (1982, Part II) provide some examples of
this. Such sources of variation need to be carefully
considered in life test work; for example, large inter-
test variation may cast test results in a less favorable
light but might also temper our reaction to large
observed differences in lab and field data.

4. There has been a very strong focus on lifetime,
or time to failure, in reliability work, but it may some-
times be more fruitful to examine other variables. For
example, in some situations it may be possible to
model physical degradation of components or equip-
ment. This might be done through the use of discrete
or continuous variables that measure degradation of
the equipment over time; the choice of suitable sto-
chastic process models in this connection is a major
problem. There has been a considerable amount of
modeling of this kind, for example in the fatigue fail-
ure literature, but there has been very little interplay
between this and statistical methodology. In a statis-
tical vein, Hooper and Amster (1982) present some
recent work. Some of the ideas involved in multistate
processes applied to biomedical problems (see, e.g.,
Kalbfleisch, Krewski, and van Ryzin 1983) may also
be useful here. See also Comment 7 in this list.

5. Robustness and model dependence should be
more routinely examined in reliability work. For ex-
ample, it is common to want to estimate a low quan-
tile of a life distribution, or a probability that is close
to zero or one. This is frequently done using a para-
metric model such as the Weibull or lognormal distri-
bution. The key problem is that, with the amount of
data on hand, there are usually different models that
fit the data adequately but that give disparate esti-
mates of tail probabilities or quantiles. For example,
Lawless (1980) considers 23 observations on the en-
durance of ball bearings. Both the Weibull and log-
normal models fit the data well, but .95 confidence
intervals for the .01 quantile under the two models are
(in millions of revolutions to failure)

\[
\begin{align*}
(3.63, 15.03) & \quad \text{Weibull} \\
(11.47, 24.53) & \quad \text{lognormal}
\end{align*}
\]

The disparity in the two results is obvious. Hahn and
Meeker (1982, Part II) discuss similar problems. There
is a need both for methods of assessing the depen-
dence of inferences on model assumptions, and for
assessing the real precision with which a quantity can
be estimated, taking into account uncertainty about
the model. Ideas of Easterling (1976) and work on
Bayesian nonparametric methods (e.g., Kalbfleisch
and Prentice 1980, Sec. 8.4) may be relevant here.

6. There is also a need for model-sensitivity analy-
sis of a slightly different sort, for example in con-
nection with risk analysis. In particular, in risk analysis,
data from a variety of sources are often pooled, and
meager data are buttressed by expert opinion, educa-
ted guesses, and so on. In such situations it is particu-
larly important to be able to assess the credibility of
any quantitative conclusions by a careful examination
of assumptions and data. For example, Bayesian
methodology provides one way to combine prior
opinions or guesses with observed data; methods for
delineating the relative effects of the data and of the
prior guesses on the conclusions would be useful. (For
a simple but interesting example of this see Barnard's
(1961) discussion of a paper by Anscombe (1961), who
considered the estimation of small probabilities.) Stat-
isticians should similarly try to develop ways of as-
sessing the effects of nonindependence of certain types
of events in a risk analysis.

7. There should be more work on the development

TECHNOMETRICS ©, VOL. 25, NO. 4, NOVEMBER 1983
of models for repeated failure analysis and multiple-event problems, and on statistical techniques for use with such models. A few comments on this area were made in Section 3.5. Specific problems of importance include (a) the design of experiments to discriminate among alternative models, and (b) information analyses that examine the extent to which a model can be tested, and its parameters estimated, from data of a certain type. There is a strong need for models and methods that break away from the traditional reliability focus on lifetime distributions; Ascher and Feingold (1983) address this point at length and discuss the application of point process models to failure and repair problems. They emphasize that statistical methodology for repairable systems has been greatly neglected by reliability workers.

6. CONCLUSION

Statistics has been successfully applied to reliability problems, and reliability has in turn stimulated developments in statistical methodology. However, there are difficulties inherent in the field of reliability that pose problems for the application of statistical methods. One is that whereas reliability problems generally provide rich opportunities for mathematical (including probabilistic) modeling, it is often difficult to obtain data capable of providing clear answers to questions of interest. In some situations, for example with systems designed to have very high reliability, statistical methods may even be of little use. In addition, the physical and human processes underlying many reliability problems are either so poorly understood or so inherently unpredictable as to make mathematical modeling merely speculative. In this setting, it is not surprising that statistics is often misused, or even ignored. It is also not surprising that there is often a wide gap between the mathematical models proposed for a situation and the ability of experimental or observational data to assess or use the models.

To me, this underlines the need for good statistical work in reliability. The problems of modeling, data collection, and data analysis are difficult. They are sufficiently hard, and each problem sufficiently unique, that I have only talked about reliability in this paper in very general terms. I have been somewhat more specific about statistical methodology that is useful in reliability. I hope that this specificity, along with the reader’s everyday familiarity with reliability (which, after all, affects most areas of human endeavor to some degree), provides some flavor of the types of problems arising in the real world.

ACKNOWLEDGMENTS

Several friends and colleagues provided me with valuable insights and comments while I was working on this paper. I would particularly like to thank Harry Ascher, Bob Little, Nancy Mann, John McCool, Bill Meeker, Ray Schafer, and an Associate Editor.

[Received December 1982. Revised March 1983.]

REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES

1. INTRODUCTION

Dr. Lawless has done an outstanding job of giving us perspective on the history, the present state, and the prospects for the future of statistical methods for reliability. In this discussion I elaborate on some of his ideas and indicate what I believe to be the areas in which we, as statisticians, can make further contributions to this field.

2. APPLICATIONS OF STATISTICS IN RELIABILITY

Statistical methods for life data analysis are used to measure, compare, and predict characteristics of the distribution of the time to some particular event or events of interest. The primary characteristic that sets life data analysis apart from other applications in statistics is that life data are usually censored, some or all exact times to failure being unobservable. This occurs, for example, because it is often not possible to wait for all units to fail or because failures are only discovered at a few inspection times. Additional complications include (a) multiple causes of failure (particularly those that interact with each other), (b) time-varying covariates, (c) the occurrence of repeated events on the same subject, and (d) the fact that the normal distribution does not play a central role as a model for life data.

There are many reasons for collecting life data. A general problem facing reliability engineers is that of predicting, estimating, and improving the reliability of products with electronic, mechanical, and chemical components. This can be a difficult problem, as reliability prediction usually involves some combination of extrapolations from

- accelerated test conditions to use conditions,
- component and subsystem reliability information to complete system behavior, or
- laboratory life test performance to field performance.