Survival, hazard and competing risks

Jayant V. Deshpande* and Sudha G. Purohit
Department of Statistics, University of Pune, Pune 411 007, India

In this paper we review certain topics in the area of survival analysis which deals with statistical/mathematical modelling of the survival/death phenomenon and its ramifications. We also discuss the major statistical procedures for suggesting and validating the models for the time to failure (lifetime) of a unit. The main concerns in survival analysis include the various censoring schemes which arise in collection of data on survival time of a unit. We study data analysis under three schemes. We study the probability laws which govern the distribution of time to failure, the effects of various covariates on these laws, as well as the estimation and comparison of effectiveness of the various competing risks.

SURVIVAL analysis is the branch of statistics which deals with collection of data, modelling and statistical analysis of data on lifetimes of units. We assume that the data consist of realizations of independent and identically distributed (i.i.d.) random variables (r.vs). Since we deal with lifetimes here, we further assume that the random variables are positive-valued and for the sake of simple modelling, continuous. Thus, we may denote the lifetime of a unit by X, which is a positive-valued continuous random variable. Let

$$F(x) = P[X \le x] \quad \text{and} \quad \bar{F}(x) = P[X > x],$$

be its cumulative distribution function (c.d.f.) and survival function, respectively. Let us assume the existence of corresponding probability density function (c.p.d.f.),

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x),$$

and the hazard rate

$$r_F(x) = \left[\frac{f(x)}{\bar{F}(x)}\right].$$

It is well known that

$$P[a < x \le b] = \int_a^b f(x)\,\mathrm{d}x,$$

gives the probability that death will occur in the interval $(a, b]$. However, the conditional probability that death will occur in the interval $(a, b]$ given that the unit has survived up to a is denoted by $P[a < X \le b \mid X > a]$ and is given by:

$$P[a < X \le b \mid X > a] = \frac{P[a < X \le b]}{P[X > a]} = \frac{F(b) - F(a)}{\bar{F}(a)}.$$

Thus, it is seen that the hazard rate ($r_F(x)$) is given by:

$$r_F(x) = \lim_{t \to 0} \frac{1}{t} P[x < X \le x + t \mid X > x],$$

and is interpreted as the (limiting) instantaneous conditional rate of failure at age x, given survival up to age x. F, $\bar{F}$, f and r_F are in one-one correspondence with each other since

$$\bar{F}(x) = \exp\left\{-\int_0^x r_F(t)\,\mathrm{d}t\right\}.$$

Since survival experiments are often conducted on live subjects (often humans), certain complications get introduced at the data collection stage itself. The main cause for this is that many subjects fail to continue to be in the study till the event of death/failure. This leads to incomplete data due to censored observations. In this article, we discuss the various censoring types and define the likelihood for the data realized from them.

Next we see that the exponential distribution plays the central role in modelling the probability laws of survival data. Certain other parametric models are also introduced to handle departures from exponentiality.

We also discuss the development due to Cox[1], who studies the effects of covariates on lifetimes through the proportional hazard model. This model is relatively easy to understand. Also statistical inferences within this model can be carried out with very few assumptions in the semiparmetric framework.

Lastly, we consider the competing risks scenario. This arises when death can be attributed to one of the several risks, unambiguously. Then we can separate the probabilities (relative failure rate, etc.)of death due to various risks, estimate those and make comparisons among them. Here also one may write the likelihood of the data as observed within this scheme and develop inference procedures.

It may not be out of place to point out that the methods of survival analysis can be applied in many areas beyond survival analysis (and reliability which is concerned with failure of industrial objects). It needs to be recognized that the data to be analysed should be in the form of the

*For correspondence.

time of occurrence of the event of interest. Such events may be from economics–time spent in the state of unemployment, from sociology–time spent out of jail until next conviction, and from other disciplines. In general, one may call this methodology as event time analysis.

Censored data

In survival analysis the observations are lifetimes which can be indefinitely long. So quite often the experiment is so designed that the time required for collecting the data is reduced. Two types of censoring are built into the design of the experiment to reduce the time taken for completing the study.

Type I (time censoring)

A number, say n, of identical units are simultaneously put into operation. However, the study is discontinued at a predetermined time t_0. Suppose n_u items have failed by this time and the remaining $n_c = n - n_u$ items remain operative. These are called the censored items. Let U be the set of ordered uncensored lifetimes $t_{\langle 1\rangle}$, $t_{\langle 2\rangle}$, . . ., $t_{\langle n_u\rangle}$, and C be the set of censored items with censoring time t_0 for the n_c censored items. In this case, the likelihood ($L(\underline{\theta})$) is:

$$L(\underline{\theta}) = \prod_{i\in U} f_{\theta}(t_{\langle i\rangle}) \prod_{i\in C} (\overline{F}_{\theta}(t_0), \quad 0 < t_{\langle 1\rangle} < \ldots < t_{\langle n_u\rangle}.$$

Type II (order censoring)

Again suppose that n identical units are simultaneously put into operation. The study is discontinued when a predetermined number, say $k(< n)$ items fail. Hence the failure times of k failed items are available. These are the k smallest-order statistics $t_{\langle i\rangle}$, $i = 1, \ldots, k$. For the remaining $(n - k)$ items the censoring time $t_{\langle k\rangle}$ which is the failure time of the item failing last is available. The likelihood of the data is given by:

$$L(\underline{\theta}) = \frac{n!}{(n-k)!} \prod_{i=1}^{k} f_{\theta}(t_{\langle i\rangle})[\overline{F}(t_{\langle k\rangle})]^{n-k}, \; 0 < t_{\langle 1\rangle} \ldots < t_{\langle k\rangle} < \infty.$$

Right random censoring

The above types of designed censoring are more prevalent in the reliability studies (of engineering systems). In survival studies (regarding biomedical subjects) censoring is more a part of the experimental situation than a matter of deliberate design. Such undesigned censoring occurs when some information about individual survival time is available, but not about exact survival time. As a simple example of such undesigned censoring consider leukaemia patients who are followed from the start of the remission until they go out of remission. The event defining failure is 'going out of remission'. If for a given patient the study ends while the patient is still in remission (i.e. the event defining failure does not occur yet), then the patient's survival time is considered as censored. For this person, it is only known that the survival time is not less than the period for which the person was observed. It is called right random censoring and is the most frequent type of random censoring.

Let $X_1, X_2, \ldots, X_n$, be the i.i.d. r.v.s. with c.d.f.F representing lifetimes of n independent, identical units on test. However, associated with each X_i, there is a random variable C_i, known as its censoring variable. We observe:

$$T_i = \min(X_i, C_i), i = 1, 2, \ldots, n, \quad \text{and}$$

$$\delta_i = \begin{cases} 1 & \text{if } X_i \le C_i, \\ 0 & \text{if } X_i > C_i. \end{cases}$$

Let the censoring variable have the p.d.f. 9 and distribution function G. It is assumed that T_i and C_i are independent. (Without this assumption, only few results are available.) One should carefully see whether the assumption of independence of T_i and C_i is justifiable. For example, in clinical trials when the reason for withdrawal is related to the course of the disease, this assumption may not be satisfied.

At the end of the study we have a sample consisting of n pairs of observations (t_i, δ_i). The likelihood is given by:

$$L(\underline{\theta}) = \prod_{t_i\in U} f_{\underline{\theta}}(t_i) \prod_{t_i\in C} \overline{F}_{\underline{\theta}}(t_i) \prod_{t_i\in U} \overline{G}(t_i) \prod_{t_i\in C} g(t_i),$$

where U is the set of uncensored (complete) observations and C is the set of censored observations.

The last two terms of the likelihood are ignored as they do not involve the unknown lifetime parameters.

A well-known non-parametric estimator of the survival function $\overline{F}$ in the absence of censoring is the empirical survival function ($\overline{F}_n(t)$) defined as follows:

$$\overline{F}_n(t) = \frac{1}{n}\begin{pmatrix} \text{number of subjects with} \\ \text{survival time longer than } t \end{pmatrix}.$$

Its generalization to randomly-censored data is the popular product limit estimator of Kaplan–Meier[2]. Let $t_1 < t_2 < \ldots < t_k$ be the distinct failure-censoring epochs in a random sample of size n ($k \le n$). Let n_i be the number of subjects having lifetime at least t_i and d_i be the number of deaths at t_i ($d_i = 0$, if only censoring takes place at t_i). Then:

$$\overline{F}(t) = \prod_{t_i\le t}\left(1 - \frac{d_i}{n_i}\right),$$

is the Kaplan–Meier (product limit) estimator of $\bar{F}(t)$. The estimator of its asymptotic variance is given by the following formula, originally developed by Greenwood[3] in another context.

$$\hat{\mathrm{Var}}(\hat{F}_{\langle t\rangle}) = (\hat{F}_{\langle t\rangle})^2 \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)}.$$

One can then obtain confidence intervals for the unknown d.f. Hall and Wellner[4] have constructed confidence bands for the entire survival function.

Parametric models

The useful concept in modelling life distributions is that of ageing. The age of a working unit (or living individual) is the time for which it is already working satisfactorily without failure.

Let a unit be of age t. It has residual lifetime T_t with $\bar{F}_t$ as its survival function. 'No ageing' phenomenon can be described as 'age has no effect on the residual lifetime of the unit'. A mathematical way of describing this would be to say that T_t $(t \ge 0)$ are identically distributed random variables, i.e.

$$\bar{F}(x) = \bar{F}_t(x) \quad \forall t,\ x \ge 0.$$

This gives, $\bar{F}(x)\bar{F}(t) = \bar{F}(t + x)$. The last equation is the celebrated Cauchy functional equation. It is well known that the exponential is the only continuous distribution that satisfies it. It has the density, $f(x, \lambda) = \lambda e^{-\lambda x}$, $x > 0$, with the corresponding hazard rate $r_F(x) = \lambda$ (constant) and $\bar{F}(x) = e^{-\lambda x}$, $x > 0$. It is seen that the survival function of T_t at age t is:

$$\bar{F}_t(x) = \frac{\bar{F}(t+x)}{\bar{F}(t)} = e^{-\lambda x},$$

which is again the same exponential distribution. We shall call this characteristic property as the 'no ageing property'. Electronic items, light bulbs, etc often exhibit the 'no ageing' phenomenon. These items do not change properties with usage; but they fail when some external shock, like a surge of high voltage comes along. Hence exponential distribution is a good model for lifetimes of such items. Davis[5] gives a number of examples, including bank statement and ledger errors, payroll cheque errors, automatic calculating machine failure and radar set component failure, in which failure data are well described by the exponential distribution. Applications in animal and human studies of chronic and infectious diseases can be found in Zelen[6], Feigl and Zelen[7], and Zippin and Armitage[8]. Using the notation and likelihood expressions from the section on 'Censored data', we state here the estimators of λ under the three censoring types:

1. Type I censoring: $\hat{\lambda} = \dfrac{n_u}{\sum_{i=1}^{n_u} t_{\langle i\rangle} + t_0(n - n_u)}$.

2. Type II censoring: $\hat{\lambda} = \dfrac{k}{\sum_{i=1}^{k} t_{\langle i\rangle} + (n-k)t_{\langle k\rangle}}$.

3. Random censoring: $\hat{\lambda} = \dfrac{n_u}{\sum_{i=1}^{n} t_i}$.

The asymptotic variances of these estimators can be obtained through the Fisher information formulae, leading to asymptotic confidence intervals.

Though there are plenty of 'no ageing' situations, we more often come across situations in which positive ageing phenomenon is observed. By positive ageing we mean that the age has adverse effect on the residual lifetime of the unit. If age affects the performance of the unit adversely, then the residual lifetime of unit of age t_2 is stochastically shorter than residual lifetime of a unit of age t_1 $(t_1 < t_2)$.

This could be stated as: $X_{t_1} \ge X_{t_2}$ $\forall\ 0 \le t_1 \le t_2$ or equivalently, in terms of survival functions:

$$\bar{F}_{t_1}(x) \ge \bar{F}_{t_2}(x) \quad \forall\ x \text{ and } 0 \le t_1 \le t_2.$$

This is equivalent to $r_F(t) \uparrow t$, if the failure rate exists. Hence the shape of the hazard function indicates how an item ages. The exponential distribution is characterized by constant hazard rate, as there is no ageing or wearout. The hazard function being increasing means that items are more likely to fail as age increases. In other words, items wear out or degrade with time. This certainly is the case with mechanical items that undergo wear or fatigue. It can also be seen in case of biomedical experiments. If T is time until a tumour appears after the carcinogenic injection in an animal experiment, then the carcinogen makes the tumour more likely to appear as time passes. The class of probability distributions of such random variables is known as 'increasing failure rate (IFR)' class. The decreasing hazard is (decreasing failure rate, DFR) less commonly observed in real life. In this case, the items are less likely to fail as time passes. Some metals work-harden through use and thus have increased strength as time passes. We shall discuss two families of distributions which are commonly used to model lifetimes of the items/individuals with increasing (decreasing) failure rates.

The Weibull family

The survival function for this family is given by: $\bar{F}(t) = e^{-\lambda t^{\gamma}}$; $t > 0$; $\lambda, \gamma > 0$. This distribution is a generalization of

the exponential distribution. However, it has a hazard rate which may have different shapes. For $\gamma = 1$, the distribution has constant hazard (i.e. exponential distribution); for $\gamma > 1$, it belongs to IFR class and $\gamma < 1$, to DFR class.

The Gamma distribution

$$f(t) = \frac{\lambda^{\gamma} t^{\gamma-1} e^{-\lambda t}}{r^{(\gamma)}}; \quad t > 0, \gamma > 0, \lambda > 0.$$

This distribution is also a generalization of the exponential distribution and includes chi-square distribution as a special case. Again $\gamma = 1$ gives the exponential distribution, for $\gamma > 1$, the IFR class and $\gamma < 1$, the DFR class. Furthermore, for all γ,

$$\lim_{t\to\infty} r(t) = \lambda,$$

indicating that a lifetime with a gamma distribution will have an exponential tail. Thus, if an item survives long enough, the distribution of the remaining lifetime is approximately exponential.

Although the exponential, Weibull and gamma distributions are popular models, several other distributions are also useful in modelling lifetimes. We mention few such models below:

Linear failure rate family

The survival function for this family is:

$$\bar{F}(t) = \exp[-(\lambda t + \tfrac{1}{2}\sqrt{\lambda t^2})], \quad \lambda, \gamma > 0, t \geq 0.$$

Makeham family

The failure rate (hazard rate) for this distribution is given by:

$$r_F(x) = [1 + \theta(1 - e^{-x})], \quad \theta \geq 0.$$

Pareto distribution

Harris[9] has pointed out that a two-parameter, Pareto distribution of the second kind (referred to as Lomax distribution) arises as a compound exponential distribution when the parameter of the exponential distribution is itself distributed as a gamma variate. For this distribution the hazard rate is

$$r_F(x) = \frac{\alpha\beta}{(\beta x + 1)}, \quad \alpha, \beta > 0, x > 0.$$

This distribution belongs to DFR class of distributions.

Non-monotonic hazard functions

Two-component parallel system: Suppose that the two components are independent and have respective life distributions:

$$F_1(t) = 1 - e^{-\lambda_1 t}, \quad \lambda_1 > 0, t \geq 0, \text{ and}$$

$$F_2(t) = 1 - e^{-\lambda_2 t}, \quad \lambda_2 > 0, t \geq 0.$$

If F is the life distribution of the system, then

$$r_F(t) = \frac{\lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2) e^{-(\lambda_1+\lambda_2)t}}{e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}}.$$

It can be verified that $r(t) \uparrow t$ on (θ, t_0) and decreases on (t_0, ∞), where t_0 depends on λ_1 and λ_2.

The system considered in the above example is a coherent system[10]. The lifetime distribution of the system belongs to what is known as increasing failure rate average (IFRA) class of life distribution. It is defined as the class of distributions for which $\bar{R}(t)$, defined as:

$$\bar{R}(t) = \frac{1}{t}\left[\int_0^t r_F(x)\, dx\right],$$

is an increasing function of t.

Another class of life distributions which arises naturally in human mortality studies and in reliability situations is characterized by failure rate function having 'bath-tub shape'. A good example of this is seen in standard human mortality tables. The risk of death for infants decreases as age advances. Then for ages 10–30, the death rate is almost constant; the deaths in this period are mainly due to random causes such as accidents. Finally, after the age of 30 an increasing proportion of the living persons die as age advances. The three phases are represented by a bath-tub curve (see Figure 1).

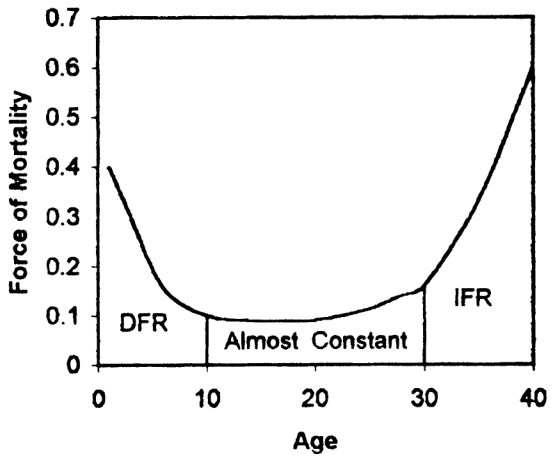

Figure 1. Bath-tub curve.

Model selection and validation

Exponential distribution is the most popular distribution for modelling lifetimes. However, exponential distribution should be used judiciously since its 'no ageing' property actually restricts its applicability. So testing for exponentiality is important. Hahn and Shapiro[11] have suggested a two-sided test for the one-parameter (unknown) exponential distribution. The test is based on W_E statistics as defined below:

$$W_E = \frac{\sum_i^n (t_i - \bar{t})^2}{\left(\sum_i^n t_i\right)^2},$$

where $t_1, t_2, \ldots, t_n$ are observed lifetimes of the n units on test. The null hypothesis of exponentiality is rejected if W_E is too small or too large.

It is seen that the exponential distribution is a member of every ageing class. The phenomenon of strictly positive (negative) ageing occurs when there is departure from the exponential distribution. So a test for exponentiality within a positive or negative ageing class when the H_0 is rejected, will suggest the ageing class in which the search for the model should be limited. Again we shall restrict only to three tests among the several tests of this type from the published literature.

Hollander and Proschan test I[12]:

H_0: $F(x) = 1 - e^{-\lambda x}$, $x \geq 0, \lambda > 0$ (unspecified).

Using Cauchy functional equation an equivalent formulation of H_0 is:

H_0: $\bar{F}(s + t) = \bar{F}(s)\bar{F}(t) \quad \forall\, s, t > 0$ vs

H_1: $\bar{F}(s + t) < \bar{F}(s)\bar{F}(t)$ for some $s, t > 0$,

(i.e. the new better-than-used property).

The test is based on the measure τ defined as:

$$\tau = \int_0^\infty \int_0^\infty \{\bar{F}(s)\bar{F}(t) - \bar{F}(s+t)\}\, dF(s)\, dF(t).$$

$$\tau = \tfrac{1}{4} - \gamma \quad \text{on simplification,}$$

where

$$\gamma = \int_0^\infty \int_0^\infty \{\bar{F}(s+t)\, dF(s)\, dF(t).$$

Here the alternative corresponds to the positive values of τ or small values of γ. So the U-statistics[13] is constructed for γ.

Let $T_1, T_2, \ldots, T_n$ be a random sample from distribution F and $T_{\langle i\rangle}$ be the ith order statistics. Define:

$$\psi(T_1, T_2, T_3) = \begin{cases} 1 & \text{if } T_1 > T_2 + T_3, \\ 0 & \text{otherwise.} \end{cases}$$

Then:

$$U = \frac{2}{n(n-1)(n-2)} \sum_{i<} \sum_{j<} \sum_{k=1}^{n} \psi(T_{\langle k\rangle}, T_{\langle j\rangle}, T_{\langle i\rangle}).$$

For large samples the test statistics is asymptotically normal. For small samples, the cut-off points are provided using Monte Carlo simulations.

Hollander and Proschan test II[14]: This test uses properties of total time on test (TTT) transform defined as:

$$\psi_F(t) = \frac{1}{\mu} \int_0^{F_{(t)}^{-1}} \bar{F}(u)\, du,$$

where μ is the mean of the distribution. Exponentiality is tested within the 'new better-than-used in expectation' class. Let $T_{\langle i\rangle}$, $i = 1, 2, \ldots, n$, be order statistics of a random sample $\{T_i, i = 1, 2, \ldots, n\}$ from the distribution F. Let

$$U_j = \sum_{k=1}^{j} \frac{(n-k+1)(T_{\langle k\rangle} - T_{\langle k-1\rangle})}{n\bar{T}_n},$$

where $\bar{T}_n$ is the sample mean. Define

$$K_n = \sum_{j=1}^{n} (U_j - j/n).$$

The test statistics is

$$\tau_n = (K_n - \tfrac{1}{2}) \frac{\bar{T}_n}{n}.$$

The test statistics is asymptotically normal and small sample cut-off points are provided by simulations.

Deshpande test[15]: This is a class of tests for exponentially within the IFRA class of life distributions. The test is based on the following characterization of IFRA property:

$$F \text{ is IFRA} \Leftrightarrow [\bar{F}(x)]^b \leq \bar{F}(bx),\ 0 \leq b \leq 1,\ x \geq 0.$$

Thus a measure of divergence from exponentiality is given as:

$$\tau = \left(\int_0^\infty \bar{F}(bx)\, dx\, (x) - \frac{1}{b+1} \right)$$

$$= \left(\gamma - \frac{1}{b+1} \right).$$

The U-statistics estimator J_b is constructed with the parameter γ. Let $T_1, T_2, \ldots, T_n$ be a random sample from distribution F. Define:

$$\psi(X_1 - bX_2) = \begin{cases} 1 & \text{if } X_1 > bX_2, \quad 0 < b < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then the test statistics is:

$$J_b = \frac{1}{n(n-1)} \sum_{i<} \sum_{j=1}^{n} [\psi(X_i - bX_j) + \psi(X_j - bX_i)].$$

The asymptotic distribution of J_b is normal. For small samples, the cut-off points are provided by Monte Carlo simulations for $b = 0.5$ and $b = 0.9$.

Graphical methods have long been used to judge the goodness of fit. These methods are simple and effective. A number of authors have discussed the use of probability plots for informal evaluation of distributional assumptions and for the estimation of parameters[16–19].

Cox's proportional hazards model: A method of regression

Our primary interest in survival analysis is the study of lifetimes. In clinical and other experimental enquiries, measurements on characteristics which possibly have influence on lifetime are also obtained. Such characteristics are called concomitant or explanatory variables or simple covariates. For example, many medical charts contain a large number of patient characteristics, i.e. values of covariates which are possibly related to the prognosis. A statistical analysis is useful in sorting out the ones that are most closely related to the prognosis. This can be done by introducing models which represent the influence of concomitant variables. Such models, which deal with the relationship between two variables, a dependent or response variable Y and an independent variable or covariate X, are known as regression models.

Model formulation

Suppose for every individual there is defined a $qx1$ vector Z of covariates. It is often convenient to define Z such that $Z = 0$ corresponds to some meaningful 'standard' condition.

The baseline lifetime model (when $Z = 0$) could be parametric or non parametric and the representation of changes introduced by $Z \neq 0$ is usually parametric. Such models are called 'semi-parametric' if baseline is not known.

Proportional hazards model (PH model): This is the most cited model which was introduced by Cox[1] in his path-breaking paper. The simple form of the proportional hazards model is:

$$h(t, Z) = h_0(t)\psi(Z),$$

where $h_0(t)$ is the baseline failure rate and $\psi(Z)$ is a parametric link function bringing in the covariates. It satisfies $\psi(0) = 1$ and $\psi(Z) \geq 0$ for all Z. The commonly used form of ψ is:

$$\psi(Z) = \psi(Z, \underline{\beta}) = \exp(\underline{\beta}'Z),$$

known as the log linear form. Thus, for the individual with covariate vector Z, the hazard function $h(t, Z)$ can be represented as:

$$h(t, Z) = h_0(t) \exp[\underline{\beta}'Z],$$

so that the ratio:

$$\frac{h(t, \underline{Z})}{h_0(t)} = \exp[\underline{\beta}'Z],$$

represents the 'relative risk' of failure. Further,

$$\log\left[\frac{h(t, \underline{Z})}{h_0(t)}\right] = \underline{\beta}'Z,$$

is the usual form of a linear regression model. Hence the name 'log linear model'. The regression coefficients are constants and covariates are fixed. Therefore, hazards $h(t, Z)$ and $h_0(t)$ are proportional, hence the name proportional hazards.

We first consider the case of single binary covariate

$$Z_i = \begin{cases} 0 & \text{if the patient is not exposed,} \\ 1 & \text{if the patient is exposed.} \end{cases}$$

Here the term exposed may refer to a risk factor such as smoking, or a patient's characteristic such as race (white, non white) or sex (male/female). It can be seen that:

$$h(t;\ \text{non exposed}) = h_0(t),$$
$$h(t;\ \text{exposed}) = h_0(t)e^{\beta},$$

so that hazard ratio, e^{β}, represents the constant relative risk of exposure. Let $\bar{F}_0$ be the survival function for non exposed and $\bar{F}_1$ be the survival function for exposed. Then

$$\bar{F}_1(t) = (\bar{F}_0(t))e^{\beta}.$$

Thus, testing for regression coefficient β is zero, is equivalent to testing equality of survival functions of non exposed and exposed.

Semi-parametric PH model: We shall consider the baseline ($h_0(t)$) as completely unknown and the covariates as fixed. Let $t_1 < t_2, \ldots, < t_m$ be the lifetimes of failed subjects. The rest of the $(n - m)$ subjects are randomly censored. For each known time of death/failure, t_j, there exists a risk set R_j of subjects which

are still in the study just prior to this time point. Cox[1] suggested using:

$$L = \prod_{j=1}^{m} \left[\frac{\exp(\underline{\beta}' Z_j)}{\sum_{k \in R_j} \exp(\underline{\beta}' Z_j)} \right],$$

as a partial or quasi likelihood function.

Efron[20] has shown that inferences about the set of regression coefficients based on Cox's partial likelihood are 'asymptotically equivalent to those based on all data'. Tsiasis[21] gives a proof of the asymptotic normality of $\hat{\beta}$. The concept of stratification is introduced by Kalbfleisch and Prentice[22], to accommodate the non-proportional case as well. However, sound mathematical basis for treating partial likelihood as ordinary likelihood is provided by Andersen and Gill[23]. This development extends the proportional hazards model to accommodate the time-dependent covariates. Following are examples of time-dependent covariates:

(i) Suppose that a treatment is applied at time $t_0 > 0$. Then one can incorporate a time-dependent binary covariate ($Z(t)$) defined as:

$$Z(t) = \begin{cases} 0 & \text{if } t < t_0, \\ 1 & \text{if } t \geq t_0. \end{cases}$$

(ii) In some industrial applications, a time-varying stress may be applied. So the covariate process will be the entire history of the stress process.

Thus, one is justified in employing the log likelihood equation:

$$\log L = \sum_{j=1}^{m} \left\{ \underline{\beta}' \underline{Z}_j - \log \left[\sum_{k \in R_j} \exp(\underline{\beta}' \underline{Z}_k) \right] \right\},$$

in the usual way to estimate the regression parameters and to test the hypothesis regarding the regression parameters.

Model validation

Treatment of time-dependent covariates leads to a simple test of goodness-of-fit for PH model with fixed covariates. Suppose X_1 is the (only) fixed covariate in the model. We define an additional covariate X_2 as $X_2 = X_1 t$ and consider the expanded model:

$$h(t; X_1 = x_1) = \lambda_0(t) e^{(\beta_1 x_1 + \beta_2 x_2)}$$

$$= \lambda_0(t) e^{(\beta_1 x_1 + \beta_2 x_1 t)},$$

and examine the significance of the hypothesis $\beta_2 = 0$. Acceptance of the hypothesis implies a goodness-of-fit of PH model for the factor X_1.

In addition to the above simple suggestion by Cox, other more complicated methods are suggested in the literature. For example, Schoenfeld[24] suggested generalization of Pearson's chi-square test when the regression parameter is assumed to be known and when it is estimated on the basis of partial likelihood. Anderson[25] has suggested a chi-square goodness-of-fit test based on the partition of the time axis. Wei[26], Gill and Schumacher[27] also have contributed to the literature on the goodness-of-fit testing of the PH model. Lin and Wei[28] have suggested a goodness-of-fit test for the PH model which is an extension of White[29] goodness-of-fit test in the ordinary likelihood setting. In their test, the measure of divergence from the PH model is taken as the difference between two model-based consistent estimators of Fisher's information matrix of the maximum partial likelihood estimator $\hat{\beta}$ in Cox's model. This difference is asymptotically normal with mean zero. Covariance matrix of this difference can be consistently estimated. Wald statistics based on this asymptotic distribution is suggested as the test statistics. Deshpande and Sengupta[30] have suggested a goodness-of-fit test for the PH model, where the alternative considered is monotonic increasing hazard ratio. The test statistics base on U-statistics considerations is shown to be equivalent to a linear rank statistics[13]. The asymptotic distribution of the test statistics is normal.

The graphical tests of goodness-of-fit are proposed by Cox[1], Key[31], Crowley and Hu[32] and Arjas[33].

Applications: The PH model has been used in diverse areas. We shall mention only a few. It is used in a study of dialysis and kidney transplants in end-stage renal disease[34], in the Stanford Heart Transplant Programme[35], in recidivism[36–37] which considers the probability that an inmate of jail will return to prison after release, in the study of break discs on high-speed trains[38], weapons systems[39], size regulation of Atlantic Halibut[40].

Extensions of Cox's model: The PH model has been extended to smooth additive function of the risk by Hastie and Tibshirani[41], Gentleman and Crowley[42].

In the context of regression trees and other piece-wise constant regression functions, work has been done by Davis and Anderson[43] and Gray[44]. Hastie and Tibshirani[45] extend Cox's model to model smooth time-varying variable effects. Kooperberg *et al.*[46] developed the powerful hazard regression (HARE) method which uses piece-wise linear regression splines, where knots and variables are adaptively selected on the basis of outcome variables to model the hazard function. The resulting model has similarities with the multivariate adaptive regression spline (MARS) model of Friedman[42]. However, because the HARE method focuses on the estimation of the entire hazard function, instead of just relative risk regression function, the unconditional log hazard function model must also be modelled parametrically as part of the

procedure. LaBlanc and Crowley[48] have considered models similar to HARE model, except that the technique proposed uses a least squares version of MARS algorithm to adaptively select knot positions for terms in the regression model. Then, for a given set of knots, the 'standard' partial likelihood is used for estimating the coefficients. Both HARE and this method are related to the method developed by Gray[44], who has used fixed knot splines in the PH model. However, HARE and LaBlanc and Crowley methods adaptively select the knot positions. LaBlanc and Crowley also illustrated the techniques with a clinical trial data for myeloma.

Competing risks

In many situations, there are several possible risks of failure. The actual cause of failure of the individual (item) may be any one of these risks. Hence these risks are said to compete for the life of the individual. The model for lifetime in the presence of such competing risks is known as the competing risks model.

The competing risks set-up is often modelled as follows. The unit is exposed to K (≥ 2) risks of failure. One and only one of these actually claims the life and is called the cause of failure (death). It is presumed that $X_1, X_2, \ldots, X_k$ are positive-valued continuous random variables denoting the lifetime (time to failure) of the unit under the k risks, respectively. In other words, X_i, called the ith latent lifetime of the unit, represents the random lifetime of the unit when the unit is exposed to the ith risk alone. However, all the k risks act simultaneously and the actual lifetime of the ith unit is another positive-valued continuous random variable $T = \min(X_1, X_2, \ldots, X_k)$. Also, let us assume that upon failure the cause of death or the risk which actually claimed the life, becomes known. It is denoted by δ and is defined as:

$$\delta = j \text{ if } T = X_j, \quad j = 1, 2, \ldots, k.$$

Thus data available from n independent copies of the unit are:

$$(T_i, \delta_i), \quad i = 1, 2, \ldots, n.$$

It can be seen that there is correspondence between the k latent lifetimes in the competing risk set-up and the lifetimes of the k components constituting a series system.

Once such data are collected, the question of making inferences about the assumed model arises. The probability model is specified by, say, joint survival function of $X_1, X_2, \ldots, X_k$ denoted by:

$$\bar{F}(x_1, x_2, \ldots, x_k) = P[X_1 > x_1, X_2 > x_2, \ldots, X_k > x_k].$$

The functions $\bar{S}_i(t)$, $i = 1, 2, \ldots, k$ defined by

$$\bar{S}_i(t) = P[T > t, \delta = i], \quad i = 1, 2, \ldots, k,$$

are called incidence functions of the k risks, respectively. Obviously $\bar{S}_i$ can be derived from $\bar{F}$. However, the converse is not always true. Berman[49] showed that if the random variables $X_1, X_2 \ldots, X_k$ are independent with distribution functions $F_i(x)$, then there does exists a one-to-one correspondence between the joint survival function

$$\prod_{i=1}^{k} \bar{F}_i(x_i),$$

and the collection

$$\{\bar{S}_i, \quad i = 1, 2, \ldots, k\},$$

of incidence functions. Earlier Cox[50] observed that in the general dependent case, $\bar{F}$ is not identifiable from

$$\{\bar{S}_i, \quad i = 1, 2, \ldots, k\},$$

Tsiatis[51] made this clear by explicitly constructing a joint survival function,

$$\bar{G}(t_1, t_2, \ldots, t_k) = \prod_{i=1}^{k} \bar{G}_i(t_i),$$

$$= \prod_{i=1}^{k} \exp\left\{-\int_0^{t_i} s_i(x) \left(\sum_{i=1}^{k} s_i(x)\right)^{-1} \mathrm{d}x\right\},$$

where

$$s_i(x) = -\frac{\mathrm{d}}{\mathrm{d}x} \bar{S}_i(x).$$

The joint survival functions $\bar{F}$ and $\bar{G}$ lead to the same set

$$\{\bar{S}_i(t), \quad i = 1, 2, \ldots, k\},$$

of incidence functions.

There have been several studies on identifiability aspect in the parametric set-up. For example, Nadas[52] shows that bivariate normal is identifiable, Arnold and Brockett[53] show that the bivariate Makeham and two versions of Pareto distributions are identifiable. They have given explicit solutions for the parameters in the above models in terms of the incidence functions.

There is no problem of identifiability in case of independent latent lifetimes. A large number of statistical procedures have been devised to make inferences about the joint survival function $\bar{F}(x_1, x_2, \ldots, x_k)$ or equivalently in the independent case, the marginal survival functions $F_i(x)$ of latent lifetimes. See David and Moeschberger[54], Bagai and Deshpande[55] for a description of such procedures for estimation, testing and other problems, when the parametric family to which F_is belong is known. These procedures are based on the basic likelihood:

$$L(t_j; \delta_{1_j}, \delta_{2_j}, \ldots, \delta_{k_j}) = \prod_{i=1}^{k} \prod_{j=1}^{n} (f_i(t_j))^{\delta_{ij}} (\bar{F}_i(t_j))^{1-\delta_{ij}}.$$

For example, if F_i is the exponential distribution with parameters λ_i, respectively, then m.l.e. of λ_i is

$$\hat{\lambda}_i = \frac{n_i}{\sum_{j=1}^{n} t_j},$$

where t_j are the lifetimes of the units, $j = 1, 2, \ldots, n$ and n_i is the member of units failed due to the ith risk. Similarly, one can derive confidence intervals and tests for specified hypotheses.

Non parametric inference

We discuss some contributions to non parametric inference regarding independent competing risks.

It may be noted first that, in the context of reliability the latent lifetimes correspond to the lifetimes of the components that are arranged in series. So it makes sense to talk of the remaining lifetime of a component which has not failed even if the system has failed. To estimate the $\bar{F}_i$s in the non parametric set-up, the system's lifetime would be regarded as censoring variable. Then the standard Kaplan–Meier product limit estimator[2] of survival function would be used for the estimation of $\bar{F}_i$s. Suppose, there are only $k = 2$ risks operating in this system. The tests for the null hypothesis H_0: $F_1(x) = F_2(x)$ have been proposed in the literature[56]. Let us suppose that $F_1(x) = F_0(x)$, a specified distribution function and $F_2(x) = F_\theta(x)$, $\theta > 0$ belong to the same parametric family. Then Bagai *et al.*[57(a,b)] derive the locally most powerful (LMP) rank tests for H_0: $\theta = 0$ against H_1: $\theta > 0$. Here rank statistics are defined in the following way: Let $T_{\langle 1\rangle} < T_{\langle 2\rangle} < \ldots T_{\langle n\rangle}$ be the ordered lifetimes of n independent systems. Let W_j be the δ corresponding to the $T_{\langle i\rangle}$. Then a statistics of the type:

$$S = \sum_{j=1}^{n} C_j W_j,$$

where C_j for $j = 1, 2$ are some appropriate scores, may be said to be a linear rank statistics. It sums the scores corresponding to the ranks of the lifetimes which are actually the latent lifetimes ($\delta = 1$) under the first risk. It is seen that in the exponential model $\bar{F}_\theta(x) = e^{-(1+\theta)x}$, the LMP rank test is based on:

$$S_1 = \sum_{j=1}^{n} W_j = \sum_{j=1}^{n} \delta_j.$$

This is sign statistics. In the logistic model with:

$$\bar{F}_\theta(x) = (1 + e^{-(x+\theta)})^{-1},$$

the LMP rank test is based on a statistics of the form:

$$V = \sum_{j=1}^{n} (1 - b_j) W_j,$$

where

$$b_j = \frac{1}{(2n+1)} + \sum_{k=2}^{j} \frac{2n(2n-2)\ldots(2n-2k+4)}{(2n+1)(2n-1)\ldots(2n-2k+3)},$$

and

$$b_1 = \frac{1}{(2n+1)}.$$

The papers[57(a,b)] also suggest certain tests based on U-statistics on heuristic grounds. Usually equivalent forms of these statistics can be expressed as linear rank statistics. Those in the literature include:

(i) $U_1 = \sum_{i=1}^{n} (i-1) W_i,$

(ii) $U_2 = 2\Sigma(2n - 1 - i) W_i.$

The limiting distribution of each of the statistics is normal through U-statistics asymptotics. The exact null distributions are found through moment-generating function technique. In fact it has been seen that U_1 has the same null distribution as the Wilcoxon-signed rank statistics, with $(n - 1)$ replaced by n[56].

Analysis of dependent competing risks

Prentice *et al.*[59], Kalbfleisch and Prentice[59], Slud *et al.*[60], have discussed the problems in general dependent case, where neither $\bar{F}(x, y)$ nor the marginal $\bar{F}_i(x)$ are identifiable from the probability distribution of (T, δ) specified by $\bar{S}_i(t)$, $i = 1, 2$. Apart from the problems discussed in the above papers, it is seen that independence of X and Y cannot be tested on the basis of the competing risk data. It seems that only statistical questions regarding the probability distribution of (T, δ) are pertinent. Prentice *et al.*[58] have raised the questions in terms of cause-specific hazard rates:

$$\lambda_i(t) = \lim_{0<\Delta_i \to 0} \frac{1}{\Delta_i} P[t < T \le \Delta_i, \delta = 2 - i \mid T > t], \quad i = 1, 2,$$

whereas Sen[61] has concentrated on the conditional probability $\Pi_i(t) = P(\delta = 2 - i \mid T = t)$. The subsequent discussion is centred on the questions raised in terms of $S_i(t)$, $i = 1, 2$ directly and the methodology is developed for these questions.

The empirical incidence functions $\bar{S}_{in}(t)$ provide consistent estimators of the true incidence functions, where $\bar{S}_{in}(t)$ are defined as:

$$\bar{S}_{in}(t) = \frac{1}{n} \sum_{j=1}^{n} I(T_j > t, \delta_j = 2 - i), \quad i = 1, 2.$$

In Deshpande and Karia[62] the problem of Bayes' estimators of the values of:

$\bar{S}_{in}(t)$ at $0 < t_i < t_2 < \ldots, t_k < \infty$,

is discussed. It is shown that the estimators of $\bar{S}_i(t_j)$ with the constraints:

$$\bar{S}_i(t_1) > \bar{S}_i(t_2) \ldots > \bar{S}_i(t_k), \quad \text{and}$$

$$\bar{S}_1(t_k) + \bar{S}_2(t_k) \le 1,$$

can be obtained with Dirichlet prior distribution for the difference:

$$p_{ij} = \bar{S}_i(t_j - 1) - \bar{S}_i(t_j).$$

Also, the complicated posterior distributions and their features can be estimated by Markov chain Monte Carlo method. Similar study of incidence functions with *a priori* constraints $S_1(t) \le S_2(t)$ is also carried out.

Doksum and Sievers[63] have developed the theory of response functions to study the differences between distribution functions. Deshpande[64] and Aras and Deshpande[65] have suggested certain models for the incidence functions of two risks.

Model I:

$$\bar{S}_1(t) = \frac{\theta}{1-\theta} \quad \bar{S}_2(t), \quad 0 < \theta < 1.$$

This model represents the situation when T and δ are independent. It is also equivalent to the proportionality of the two cause-specific hazard rates.

Model II: $\bar{S}_1(\alpha_t) = \bar{S}_2(t), \quad t > \alpha > 0$. In the joint environment, the first risk is as potent up to age t as the second risk up to age α_t, for every $t > 0$.

Model III: $\bar{S}_1(t) = (\theta \bar{S}(t))^{\nu}$; $\bar{S}_2(t) = \bar{S}(t) - (\theta \bar{S}(t))^{\nu}$, $\nu > 0$, $0 < \theta < 1$, where $\bar{S}(t) = P[T > t]$.

Deshpande and Karia[66] have worked out response functions in the sense of Doksum and Sievers[63] and have carried out descriptive analysis for the data set of Nair[67] for these three models. In the same paper, after defining the response functions in the three models, asymptotic confidence bounds based on competing risk data have been established.

In case the two competing risks are dependent, then in order to see whether the two risks are equally potent (in the environment in which they jointly operate), it is interesting to test the hypothesis of bivariate symmetry against one-sided alternative, i.e. say:

$$H_0 : F(x, y) = F(y, x) \quad \forall\, x, y,$$

against

$$H_1 : F(x, y) > F(y, x) \quad \forall\, x > y.$$

Deshpande[64] has shown that U_1 statistics discussed earlier is useful for this testing problem.

If we define by $s_i(t)$ for $i = 1, 2$, the subdensity functions corresponding to two incidence functions, then the marginal likelihood of $R^1 = (R_1, \ldots, R_n)$, the ranks of $(T_1, \ldots, T_n)$ and the corresponding indications $W^1 = (W_1, W_2, \ldots, W_n)$ can be written as

$$P[\theta \mid R, W] = \int \prod_{i=1}^{n} \{s(t_i, \theta)\}^{W_i} \{h(t_i) - s(t_i, \theta)\}^{1-W_i} \, dt,$$

where the integration is over $t_1 < t_2 < \ldots < t_n$ and $h(t)$ is the density corresponding to $H(t) = P[T \le t]$.

Aras and Deshpande[65] show that the locally most powerful test for H_0: $\theta = \theta_0$ against H_1: $\theta > \theta_0$ is given by:

$$\text{Reject } H_0, \text{ if } L = \sum_{j=1}^{n} W_j a_j \text{ is too large.}$$

Here

$$a_j = \int \frac{s'(t_j, \theta_0)}{s(t_j, \theta_0)} \prod_{i=1}^{n} S(t_i, \theta_0) \, dt_i,$$

and the integration is over $t_1 < t_2, \ldots, < t_n$. If the subdensities s_1 and s_2 are modified from well-known probability densities, then we get familiar tests adapted to competing risks as the LMP rank tests.

Aly *et al.*[68] have used Kolmogorov–Smirnov type statistics:

$$D_n = \sqrt{n} \mid \bar{S}_{1n}(t) - \bar{S}_{2n}(t) \mid$$

for the empirical versions of incidence functions, to compare the two incidence functions. Sun and Tiwari[69] have constructed tests for equality of the two cause-specific hazards based on the difference between the Aalen estimators of two cumulative cause-specific hazard rates.

Coherent systems

In reliability theory, a well-studied class of systems is the class of coherent systems (see Barlow and Proschan[10] for details). These systems include series, parallel, k-out-of-n, etc.

There is not much work done on inference from the data realized from monitoring of coherent systems. Deshpande *et al.*[70] have considered a two-sample problem. Suppose n_1 and n_2 independent copies of a coherent system composed of k components operate in different environments. Neither the independence of the lifetimes is assumed nor is the structure of the system known. The observation scheme is the continuous monitoring model. In this situation, it is not possible to estimate either joint or marginal survival functions of the components. Hence the test developed is based on incidence functions of the k components in the two environments. Let T_{lj} and X^*_{lji} indicate the lifetime of the jth system in l_{lj}th ($l_{lj} = 1, 2$)

group and the lifetime of the ith component in the jth system, in the lth group; $i = 1, \ldots, k$ and $j = 1, 2, \ldots, n_i$: The available data are:

$$\{X_{lji} = \min(X^*_{lji}, T_{ij}),\ \delta_{lji} = I(X^*_{lji} \le T_{lj})\},$$

for all i, j and l. The incidence functions of k components in the two environments are defined as:

$$\bar{F}_{li}(x) = P[X_{lji} > x, \delta_{lji} = 1],$$

and the corresponding hazard function as:

$$\nu_{li}(x) = \frac{f_{li}(x)}{\bar{F}_{li}(x)},$$

where f_{li} is the subdensity corresponding to $\bar{F}_{li}$. Then to test the hypothesis:

$$H_0: F_{1i}(x) = F_{2i}(x\ ),\ i = 1, 2, \ldots, k,$$

the statistics proposed is of the form:

$$S = n^{-1} Z' D^{-1} Z,$$

where

$$Z' = (Z_1, Z_2, \ldots, Z_k)$$

and

$$Z_i = \int_0^T K_i(s) \left\{ \frac{d\hat{F}_{1i}(s)}{(1 - \hat{F}_{1i}(s))} - \frac{d\hat{F}_{2i}(s)}{(1 - \hat{F}_{2i}(s))} \right\},$$

where

$$\hat{F}_{li}(t) = \frac{1}{n_l} \sum_{j=1}^{n_i} I\,[X_{lji} \le t, \delta_{lji} = 1].$$

$K_i(t)$ are appropriate weight functions and T_i is large but finite, which satisfies $\bar{F}_{li}(T_i) < 1$.

Asymptotic distribution of the test statistics is chi-square, with k degrees of freedom. The matrix D in $\bar{H}$ is estimated covariance matrix which is diagonal as elements of Z are asymptotically independent. The weight functions of $\bar{K}_i(s)$ are predictable processes and different choices of these functions yield different tests of this class.

This work is a generalization of Gray[71] to consideration of more than one risk simultaneously and of Bagai and Deshpande[56] to dependent risks.

1. Cox, D. R., *J. R. Stat. Soc.*, 1972, **34**, 187–220.
2. Kaplan, E. L. and Meier, P., *J. Am. Stat. Assoc.*, 1958, **53**, 457–481.
3. Greenwood, M., Reports on Public Health and Medical Subjects, Her Majesty's Stationary Office, London, 1926, vol. 20, pp. 366–372.
4. Hall, W. J. and Wellner, J. A., *Biometrika*, 1980, **67**, 133–143.
5. Davis, D. J., *J. Am. Stat. Assoc.*, 1952, **47**, 113–150.
6. Zelen, M., *J. R. Stat. Soc. Ser. A*, 1966, **129**, 368–398.
7. Feigl, P. and Zelen, M., *Biometrics*, 1965, **21**, 826–838.
8. Zippin, C. and Armitage, P., *Biometrics*, 1966, **22**, 665–672.
9. Harris, C. M., *Oper. Res.*, 1968, **16**, 307–313.
10. Barlow, R. E. and Proschan, F., *Theory of Reliability and Life Testing*, Holt, Rinehart and Winston Inc., New York, 1975.
11. Hahn, G. J. and Shapiro, S. S., *Statistical Models in Engineering*, Wiley, New York, 1967.
12. Hollander, M. and Proschan, F., *Ann. Math. Stat.*, 1972, **43**, 1136–1146.
13. Puri, M. and Sen, P., *Nonparametric Methods in Multivariate Analysis*, John Wiley and Sons, New York, 1971.
14. Hollander, M. and Proschan, F., *Biometrika*, 1975, **62**, 585–594.
15. Deshpande, J. V., *Biometrika*, 1983, **70**, 514–518.
16. Chernoff, H. and Lieberman, G. J., *J. Am. Stat. Assoc.*, 1954, **49**, 778–785.
17. Daniel, C., *Technometrics*, 1959, **1**, 311–341.
18. Wilk, M. B., Gnanadesikan, R. and Huyett, M. J., *Technometrics*, 1962, **4**, 1–20.
19. King, J. R., *Probability and Decision Making*, Industrial Press, New York, 1971.
20. Efron, B., *J. Am. Stat. Assoc.*, 1977, **72**, 557–565.
21. Tsiatis, A. A., *Ann Stat.*, 1981, **9**, 93–108.
22. Kalbfleisch, J. D. and Prentice, R. L., *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1981.
23. Andersen, P. K. and Gill, R. D., *Ann Stat.*, 1982, **10**, 1100–1120.
24. Schoenfeld, D., *Biometrika*, 1980, **67**, 145–153.
25. Andersen, P. K., *Biometrics*, 1982, **38**, 67–77.
26. Wei, L. J., *J. Am. Stat. Assoc.*, 1984, **79**, 649–652.
27. Gill, R. and Schumacher, M., *Biometrika*, 1987, **74**, 289–300.
28. Lin, D. Y. and Wei, L-J., *Stat. Sin.*, 1991, **1**, 1–17.
29. White, H., *Econometrica*, 1982, **50**, 1–25.
30. Deshpande, J. V. and Sengupta, D., *Biometrika*, 1995, **82**, 251–261.
31. Key, R., *App. Stat.*, 1977, **26**, 227–237.
32. Crowley, J. and Hu, M., *J. Am. Stat. Assoc.*, 1977, **72**, 27–36.
33. Arjas, H., *J. Am. Stat. Assoc.*, 1989, **83**, 204–212.
34. Williams, G. W., Weller, J. M., Ferguson, W., Forsythe, S. B. and Shu-Chen Wu, *Kidney Int.*, 1983, **83**, 691–693.
35. Cox, D. R. and Oakes, D., *Analysis of Survival Data*, Chapman and Hall Ltd., London, 1984.
36. Schmitt, P. and Witte, A. D., *Predicting Recidivism Using Survival Models*, Springer-Verlag Inc., New York, 1988.
37. Ellermann *et al.*, *Oper. Res.*, 1992, **40**, 485–504.
38. Bendell *et al.*, *Qual. Reliabil. Eng. Int.*, 1986, **2**, 45–52.
39. Gray *et al.*, *Reliabil. Eng. Sys. Saf.*, 1988, **21**, 245–269.
40. Smith, S. J., Waiwood, K. G. and Neilson, J. D., in *Case Studies in Biometry* (eds Lange *et al.*), John Wiley and Sons Inc., New York, 1994, pp. 125–145.
41. Hastie, T. and Tibshirani, R., *Generalized Additive Models*, Chapman and Hall, London, 1990.
42. Gentleman, R. and Crowley, J., *Biometrics*, 1991, **47**, 1283–1296.
43. Davis, R. and Anderson, J., *Stat. Med.*, 1989, **8**, 947–962.
44. Gray, R. J., *J. Am. Stat. Assoc.*, 1992, **87**, 942–951.
45. Hastie, T. and Tibshirani, R., *J. R. Stat. Soc. Ser. B*, 1993, **55**, 757–800.

46. Kooperberg, C., Stone, C. J. and Truong, Y. K., *J. Am. Stat. Assoc.*, 1995, **90**, 78–94.
47. Friedman, J. H., *Ann Stat.*, 1991, **19**, 1–14.
48. LaBlanc, M. and Crowley, J., *Biometrics*, 1999, **55**, 204–213.
49. Berman, S. M., *Ann. Math. Stat.*, 1963, **34**, 1104–1106.
50. Cox, D. R., *J. R. Stat. Soc. Ser. B*, 1959, **21**, 411–421.
51. Tsiatis, *Proc. Acad. Sci. USA*, 1975, **72**, 20–22.
52. Nadas, A., *Technometrics*, 1971, **13**, 201–202.
53. Arnold, B. C. and Brockett, P. L., *Scand. Actuar. J.*, 1983, 117–127.
54. David, H. A. and Moeschberger, M. L., *The Theory of Competing Risks*, Monograph, Griffin, London, 1978.
55. Bagai, I. and Deshpande, J. V., Proceedings of Indian Society of Probability and Statistics, 1986, pp. 267–282.
56. Bagai, I. and Deshpande, J. V., *J. Nonparametric Stat.*, 1995, **5**, 21–31.
57. a. Bagai, I., Deshpande, J. V. and Kochar, S. C., *Commun. Stat. (Theory–Methods)*, 1989, **18**, 107–120; b. Bagai, I., Deshpande, J. V. and Kochar, S. C., *Biometrika*, 1989, **76**, 775–781.
58. Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N. and Farewell, V. T., *Biometrics*, 1978, **34**, 541–551.
59. Kalbfleisch, J. D. and Prentice, R. L., *Statistical Analysis of Failure Time Data*, Wiley, 1980.
60. Slud, E. V., Byar, D. P. and Schatztein, A., *Biometrics*, 1988, **44**, 1203–1204.
61. Sen, P. K., *Contributions to Statistics, J. Hajek Memorial Volume* (ed. Jureikova, J.), Reidel, Dordrecht, 1978.
62. Deshpande, J. V. and Karia, S. R., *IMS Lecture Notes Monograph Series*, 1995, vol. 27, pp. 79–94.
63. Doksum, K. A. and Sievers, G. L., *Biometrika*, 1976, **63**, 421–434.
64. Deshpande, J. V., *Biomet. J.*, 1990, **6**, 737–746.
65. Aras, G. and Deshpande, J. V., *Stat. Decision*, 1992, **10**, 323–336.
66. Deshpande, J. V. and Karia, S. R., *Indian Assoc. Probability, Quality Reliability Trans.*, 1995, **20**, 79–92.
67. Nair, V. N., *Int. Stat. Rev.*, 1993, **61**, 169–182.
68. Aly, E. E., Mekeague, I. W. and Kochar, S. C., *J. Am. Stat. Assoc.*, 1994, **89**, 994–999.
69. Sun, Y. and Tiwari, R. G., *Analysis of Censored Data* (eds Koul, H. L. and Deshpande, J. V.), IMS-LNMS, Hayword, 1995, vol. 27.
70. Deshpande, J. V., Dharmadhikari, A. D. and Naik-Nimbalkar, U. V., *Biometrika*, 1993, **80**, 877–883.
71. Gray, R. J., *Ann Stat.*, 1988, **16**, 1141–1154.