Competing Risks

Bo Henry Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim
Norway
bo@math.ntnu.no

8th August 2006

Abstract

Consider a unit which can experience any one of k competing failure types, and suppose that for each unit we observe the time to failure, T, and the type of failure, $C \in \{1, 2, \ldots, k\}$. The case of observing the pair (T, C) is termed "competing risks" in the statistical literature. After considering some examples we review basic notation and theory of competing risks. In particular we consider the latent failure time approach to competing risks in which the k risks are represented by potential failure times $T_1, \ldots, T_k$ where only the smallest, $T = \min_i T_i$, is observed together with its index $C = \arg\min_i T_i$. In reliability studies, the marginal distributions of the T_i are often of primary interest, but are unfortunately non-identifiable in general. Additional, though non-testable, assumptions to obtain identifiability are considered, as are bounds for the marginal distributions given in terms of observable functions. The likelihood function of right censored competing risks data is given and its consequences for both parametric and non-parametric estimation are explained. Extensions of the classical theory of competing risks to more general Markov models and to repairable systems are briefly discussed.

Keywords: Competing risks; Latent failure times; Reliability databases; subdistribution function; Cause-specific hazard function; Identifiability; Preventive maintenance; Repairable system.

1 Introduction

Suppose that units under study can experience any one of several distinct failure types, and that for each unit we observe both the time to failure and the type of failure. Failure may here, for example, correspond to breakdown of a mechanical component where there are several possible root causes for the failure, such as vibration, corrosion, etc. This is a typical case of a "competing risks" situation in reliability. The theory of competing risks does not, however, originate from reliability theory. In fact, it can be traced back to David Bernoulli's attempts in 1760 to disentangle the risk of dying from smallpox from other causes. This is a classical example of competing risks, where individuals are subject to multiple causes of death. Similar applications occur in demography and actuarial science, usually under the name of multiple-decrement analysis.

1.1 Formal definition of competing risks

Formally one observes the pair (T, C) where $T > 0$ is the time of failure and $C \in \{1, 2, \dots, k\}$ represents the type of failure. It is thus assumed that there are k different failure types, and that each failure can be classified as belonging to exactly one of the k types. Note that "failure" is used as a generic term and may in practice correspond to any event of interest depending on the application at hand. Also, "time" need not mean calendar time, but can in principle be any suitable measurement which is non-decreasing with calendar time, such as operation time, number of cycles, number of kilometers run, length of a crack etc.

An intuitive way of describing a competing risks situation with k risks, is to assume that to each risk is associated a failure time T_j, $j = 1, ..., k$. These k times are thought of as the hypothetical failure times if the other risks were not present, and they are referred to as latent failure times. When all the risks are present, the observed time to failure of the system is the smallest of these failure times along with the actual cause of failure. Thus by letting $T = \min\{T_1, ..., T_k\}$ and $C = c$ if $T = T_c$, we observe the pair (T, C) and we are back to the formulation of the previous paragraph. Note that it is assumed that the T_c for which minimum is attained is uniquely given.

1.2 Uses of competing risks in reliability and maintenance studies

Crowder [11, Ch. 1] gives some simple examples of the uses of competing risks in reliability studies. One of these is taken from King [18] who studied data of breaking strengths of certain wire connections. Two types of failure were defined (so $k = 2$): breakage at the bonded end and breakage along the wire itself.

Mendenhall and Hader [22] presented data of times to failure for VHF communication transmitter-receivers. Again two types of failures were considered: those confirmed on arrival at the maintenance center and those unconfirmed.

Modern reliability databases usually distinguish between a large number of failure modes, which suggests the use of methods from the theory of competing risks. Cooke [8, 9] reviews some main styles in the design of reliability databases, as well as models and methods for their analysis.

Failure modes are in databases often grouped into critical failures, degraded failures and incipient failures. Cooke [8] points out that whereas critical failures are of primary interest in risk and reliability calculations, a maintenance engineer is also interested in degraded and incipient failures, while a component designer may be is interested in the particular component function that is lost and in the failure mechanisms. Each of these interests leads to a different analysis, but they are all best solved by methods from competing risks.

Traditionally, competing risks were analyzed as if they were independent of each other. This assumption appears to be dubious in applications like the ones mentioned above, however. Even if assumptions of stochastic independence may often be justified by the physically independent functioning of components, a dependence between risks may be introduced by, for example, load-sharing between components or other shared common factors such as working environment, manufacture and maintenance.

A simple case of dependent risks (i.e. dependent latent times $T_1, \ldots, T_k$) occur in the case when a potential component failure at some time T_1 may be avoided by a preventive maintenance (PM) at time T_2 (see [9], [4], [21]). The assumption that T_1 and T_2 are independent is clearly unreasonable in this application, since the maintenance crew is likely to have some information regarding the component's state during operation. This insight is used to perform maintenance with the aim of avoiding component failures. Thus we are faced with a case of dependent competing risks between the variables T_1 and T_2. Note that the observable result is the pair (T, C), rather than the latent times T_1 and T_2. The latter are, however, the times of primary interest. For example, knowing the distribution of T_1, which is the true failure time distribution, could be the basis for maintenance optimization. However, as will be discussed later, in a competing risks case the marginal distributions of the T_j are not identifiable from observation of (T, C) alone, unless specific assumptions are made on the dependence between T_1 and T_2.

1.3 Basic literature on competing risks

The recent book by Crowder [11] gives a comprehensive review of the the theory and methods of competing risks. An older book devoted to the subject is David and Moeschberger [12]. Several standard books on reliability and survival analysis contain chapters on competing risks, for example Lawless [19], Kalbfleisch and Prentice [17], Nelson [25], Bedford and Cooke [4] and Andersen et al. [2]. Among several review papers written on the subject we mention Gail [14] and Moeschberger and Klein [23].

2 Model specification

The joint distribution of the pair (T, C) from an individual is completely specified by the sub-distribution functions

$$F_j(t) = P(T \leq t, C = j); \quad t > 0,\ j = 1, ...k$$

defined for $t > 0$, $j \in \{1, 2, \ldots, k\}$. In the formulas given in the following, the ranges of t and j will be as for the $F_j(t)$, and will be mostly suppressed.

The corresponding sub-density functions, when they exist, are given by differentiation,

$$f_j(t) = F_j'(t)$$

The marginal distribution of T is given by the distribution function

$$F(t) = P(T \leq t) = \sum_{j=1}^{k} F_j(t)$$

or the survival function

$$\bar{F}(t) = 1 - F(t)$$

Note that here and in the sequel, a bar above a capital letter means that this is the survival function corresponding to the distribution function given without a bar. This applies also to sub-distribution functions. Thus we define the sub-survival functions as

$$\bar{F}_j(t) = P(T > t, C = j)$$

The marginal distribution of C is given by

$$\pi_j = P(C = j) = F_j(\infty)$$

Note that then

$$F_j(t) + \bar{F}_j(t) = \pi_j$$

The distribution of (T, C) can alternatively be specified by the sub-hazard functions, which when they exist are given by

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P\left(T \leq t + \Delta t, C = j | T > t\right)}{\Delta t} = \frac{f_j(t)}{\bar{F}(t)} \tag{1}$$

It follows that

$$\lambda(t) = \sum_{j=1}^{k} \lambda_j(t) \tag{2}$$

is the hazard function of T. Moreover, from equation (1) follows that $f_j(t) = \lambda_j(t)\bar{F}(t)$ which by integration gives the useful connection

$$F_j(t) = \int_0^t \lambda_j(u)\bar{F}(u)du \tag{3}$$

Next, defining the cumulative sub-hazard functions as

$$\Lambda_j(t) = \int_0^t \lambda_j(u)du$$

it is seen from (2) that $\Lambda(t) = \sum_{j=1}^{k} \Lambda_j(t)$ is the cumulative hazard function of T. Thus we have

$$\bar{F}(t) = e^{-\Lambda(t)} = e^{-\sum_{j=1}^{k} \Lambda_j(t)} = \prod_{j=1}^{k} \bar{G}_j^*(t) \tag{4}$$

where we define

$$\bar{G}_j^*(t) = e^{-\Lambda_j(t)} \tag{5}$$

Note that $\bar{G}_j^*(t)$ is a survival function (possibly with an atom at infinity), but that it is not in general the distribution of any observable random variable. We shall see later, however, that it is the marginal distribution of T_i under the model with independent latent failure times.

The sub-hazard functions $\lambda_j(t)$ have the intuitive interpretation as the failure rate from a specific cause conditional on survival up to time t. It is also known under the names mode-specific or cause-specific hazard function, and has in older literature been called the crude hazard rate.

2.1 Latent failure time representation

Consider again the representation where $T = \min\{T_1, ..., T_k\}$ and $C = c$ if $T = T_c$ is observed, with C assumed uniquely given. Let the joint survival function of $T_1, ..., T_k$ be $\bar{K}(t_1, \ldots, t_k) = P(T_1 > t_1, \ldots, T_k > t_k)$. Then the survival function of T can be evaluated as $\bar{F}(t) = \bar{K}(t, t, \ldots, t)$. The sub-density functions can also be calculated directly from the joint survival functions as

$$f_j(t) = -\left(\frac{\partial \bar{K}(t_1, \ldots, t_k)}{\partial t_j}\right)_{t_1 = \ldots = t_k = t} \tag{6}$$

and it further follows that

$$\lambda_j(t) = \frac{f_j(t)}{S(t)} = -\left(\frac{\partial \log \bar{K}(t_1, \ldots, t_k)}{\partial t_j}\right)_{t_1 = \ldots = t_k = t} \tag{7}$$

Let the marginal survival function of T_j be denoted $\bar{G}_j(t) = P(T_j > t)$ and let the corresponding hazard rate function be $h_j(t) = -\bar{G}'_j(t)/\bar{G}_j(t)$. It is noted that in general $h_j(t)$ and $\lambda_j(t)$ are different and have different interpretations. The former has traditionally been called the net rate, while the latter is the crude rate as mentioned before. It will be seen later, however, that $h_j(t) = \lambda_j(t)$ for all $t > 0$ when the T_j are independent.

2.2 The identifiability problem

As already explained, the main interest in a competing risks analysis is often in the joint and marginal distributions of the latent failure times $T_1, ..., T_k$. The problem turns out to be, however, that the distribution of the observable pair (T, C) does not in general determine the distribution of the latent failure times. In standard terms, the joint and marginal distributions of $T_1, ..., T_k$ are non-identifiable from observation of (T, C). This means that there are several different joint distributions of $T_1, ..., T_k$ which give rise to the same distribution of (T, C). This fact was noted by Cox [10] for the case of two failure causes, while Tsiatis [28] studied the general case.

The main result of Tsiatis [28] states that if the set of sub-distribution functions $F_j(t)$ is given for some model with dependent risks, then there exists a unique model with independent risks yielding the same $F_j(t)$. This model is defined by the joint survival function $\bar{K}(t_1, \ldots, t_n) = \prod_{j=1}^{k} G_j^*(t_j)$ where the $G_j^*(t)$ are given by equation (5).

The result establishes that to each dependent competing risks model there corresponds a unique independent risks model with the same distribution of (T, C). Thus, one cannot know, from observations of (T, C) alone, which of the two models is correct, since they will both fit the data equally well.

3 How to deal with the identifiability problem

In this section we consider ways of overcoming the identifiability problem under the latent variable representation of competing risks. It should be stressed that this can only be done by imposing additional restrictions in the model. These may be of various kinds, but one should always have in mind that for given observations of the pair (T, C), the assumptions will always be non-testable.

3.1 Assuming independent risks

The classical assumption is that the risks act independently, so that the latent failure times T_j are independent. It then follows from Tsiatis [28] that we have identifiability of the distributions in question (under regularity assumptions), meaning that the marginal distributions of the T_j now can be computed from the sub-distribution functions $F_j(t)$. In practice this means that the marginal distributions can be estimated in a consistent manner from competing risks data. Furthermore, in the case of independence we can write $\bar{K}(t_1, \ldots, t_k) = \prod_{j=1}^{k} \bar{G}_j(t_j)$ and hence it follows from equation (7) that

$$\lambda_j(t) = h_j(t)$$

and further it follows that

$$\bar{G}_j^*(t) = \bar{G}_j(t)$$

where $\bar{G}_j^*(t)$ was defined in equation (5). Thus in this case the $\bar{G}_j^*(t)$ are in fact the marginal survival functions of the T_j.

3.2 Assuming a known copula for the latent variables

Zheng and Klein [31] generalized the result on identifiability in the independent risks case, proving that the marginal distributions are identifiable when the dependence is given by a known copula. Consider, as in Zheng and Klein [31], the case $k = 2$. Let K be the joint distribution function of (T_1, T_2) while G_1, G_2 are the marginal distribution functions. Then the copula of (T_1, T_2) is defined by

$$C(u_1, u_2) = K\left(G_1^{-1}(u_1), G_2^{-1}(u_2)\right); \quad (u_1, u_2) \in [0,1] \times [0,1]$$

It is well known (Nelsen [24]) that this is a joint distribution function on $[0,1] \times [0,1]$ with uniform marginals. Note in particular that independence of T_1, T_2 leads to the so called independence copula, $C(u_1, u_2) = u_1 u_2$.

Zheng and Klein [31] proved that, under certain regularity conditions, if the copula $C(\cdot, \cdot)$ is known, then the marginal distribution functions G_1, G_2 of T_1, T_2, respectively, are uniquely determined by the sub-distribution functions F_1, F_2. Thus, provided the copula is known, we are able to estimate the marginal distributions from observations of (T, C). Note that the assumption of independence can be interpreted as a case of knowing the copula, namely the independence copula as defined above. In practice the copula may not be completely known, however, but Zheng and Klein [31] suggested how partial knowledge of it can lead to bounds on the marginal survival functions.

3.3 Computing bounds for the marginal survival functions

Peterson [26] gave bounds for the joint distribution function $K(t_1, \ldots, t_k)$ and for the marginal distribution functions $G_j(t) = P(T_j \leq t)$ in terms of the observable sub-distribution functions F_j. The bounds for the marginal distribution functions are given by

$$F_j(t) \leq G_j(t) \leq F(t) \tag{8}$$

which are easily verified. Peterson [26] showed, moreover, that the bounds are pointwise sharp meaning that given the F_j, there is for each t a feasible marginal distribution function

G_j which gives equality. This applies to the lower as well as the upper bound. The bounds are not, however, functionally sharp, meaning that not every set of distribution functions G_j bounded as above can serve as marginal distributions of the T_j. Crowder [11] found the additional property that $G_j(t) - F_j(t)$ needs to be a non-decreasing non-negative function. Subsequently, Bedford and Meilijson [6] obtained the complete characterization of the feasible marginal distribution functions and showed in particular that the non-decreasingness of $G_j(t) - F_j(t)$ is also sufficient, provided a subtle additional measure theoretic assumption is satisfied. Note that when F_j and G_j are differentiable, this leads to the inequality

$$f_j(t) \leq g_j(t) \text{ for all } t > 0 \tag{9}$$

where $g_j(t)$ is the density function of T_j. We will use this inequality in the following example.

Example (adapted from Bedford and Meilijson [6]). Consider the model with $k = 2$ given by constant sub-hazard functions $\lambda_j(t) = \lambda_j$, $j = 1, 2$. In this case $F_j(t) = (\lambda_j/\lambda_+)(1 - e^{-\lambda_+ t})$ and $F(t) = 1 - e^{-\lambda_+ t}$, where $\lambda_+ = \lambda_1 + \lambda_2$. Assume now that T_1 has an exponential marginal distribution, so that $G_1(t) = 1 - e^{-\lambda t}$ for some λ. The upper bound of the inequalities (8) easily gives $\lambda \leq \lambda_+$. Further, the inequality (9) with $j = 1$ gives $\lambda_1 e^{-\lambda_+ t} \leq \lambda e^{-\lambda t}$ for all $t > 0$, which implies $\lambda \geq \lambda_1$ by letting $t \to 0$. Thus we have shown that any feasible value of λ satisfies the inequality

$$\lambda_1 \leq \lambda \leq \lambda_+$$

Bedford and Meilijson [6] in fact show that the set of possible values is the half-open interval $[\lambda_1, \lambda_1 + \lambda_2)$. Note that an assumption that T_1 and T_2 are independent leads to $\lambda = \lambda_1$. The example therefore shows that the independence assumption leads to the most optimistic value of the failure rate λ among the ones that are possible when the dependence is not specified. Williams and Lagakos [30] proved a corresponding result in a more general setting.

3.4 Parametric identifiability

Note that the meaning of identification of marginal distributions as discussed above has been in the non-parametric sense that the marginal distribution functions G_j can be derived in terms of the sub-distribution functions F_j. If a parametric model is specified for the latent failure times, then the identifiability problem is a completely different one since it now has to do with identification of a finite set of parameters. Crowder [11, Ch. 7.7] and Moeschberger and Klein [23] review models for which identification holds. Some examples of parametric models for dependent latent variables are given in the next section.

4 Modelling of competing risks

4.1 Modelling of sub-distributions

4.1.1 Mixture models

These are models given by specifying sub-distribution functions of the form

$$F_j(t) = \pi_j Q_j(t)$$

for given (parametric) distribution functions $Q_j(t)$. Typically one might let $Q_j(t)$ correspond to a Weibull distribution, so that

$$\bar{Q}_j(t) = \exp\{-\left(\frac{t}{\theta_j}\right)^{\alpha_j}\} \tag{10}$$

4.1.2 Modelling sub-hazard functions

A common approach is to assume parametric models for the sub-hazard functions, for example using Weibull hazards,

$$\lambda_j(t; \alpha_j, \theta_j) = \frac{\alpha_j}{\theta_j}\left(\frac{t}{\theta_j}\right)^{\alpha_j} \tag{11}$$

Note that this model has fewer parameters than the model (10) since the latter needs a specification of the π_j in addition to the pairs (α_j, θ_j).

4.1.3 Regression models

Let $\mathbf{x}$ be a vector of covariates for the unit under study. Two main approaches for regression modelling in survival analysis are proportional hazards modelling and accelerated life modelling. The versions for competing risks can be given as follows.

Proportional hazards

Let the sub-hazard functions be given by

$$\lambda_j(t; \mathbf{x}) = \psi_j(\mathbf{x})\lambda_{0j}(t)$$

where $\psi_j(\mathbf{x})$ is a positive function of the covariates. Usually such a function is on parametric form, for example $\psi_j(\mathbf{x}) = \exp\{\boldsymbol{\beta}'\mathbf{x}\}$ for a parameter vector $\boldsymbol{\beta}$. The $\lambda_{0j}(t)$ are called baseline sub-hazards. Sometimes one assumes that $\lambda_{0j}(t) = \lambda_0(t)$ does not depend on j, in which case T and C are stochastically independent.

Accelerated life model

The dependence of $\mathbf{x}$ is here through factors $\phi_j(\mathbf{x})$ which accelerate time in such a manner that

$$F_j(t; \mathbf{x}) = F_{0j}(\phi_j(\mathbf{x})t)$$

4.2 Modelling of latent variables

The traditional way of modelling dependent risks has been through specification of the joint distribution function $K(t_1, \ldots, t_n)$ or the joint survival function $\bar{K}(t_1, \ldots, t_n)$.

Bivariate exponential (Gumbel [15])

Let $k = 2$ and define

$$\bar{K}(t_1, t_2) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \nu t_1 t_2\}$$

Then $\lambda_j(t) = \lambda_j + \nu t$ so that the independent risks model with the same sub-distribution is given by

$$\bar{G}_j^*(t) = \exp\{-\lambda_j t - \nu t^2/2\}$$

while the marginal distribution for the given $\bar{K}$ are

$$\bar{G}_j(t) = \exp\{-\lambda_j t\}$$

Note that it is not possible to distinguish between these models from observation of (T, C).

Frailty models

A class of models, called "frailty models" (see Crowder [11, Ch. 3]) are obtained by assuming that $T_1, \ldots, T_k$ are independent given a random "frailty" variable Z which varies from unit to unit. More precisely, assume that Z is a positive random variable and that conditionally given $Z = z$ the $T_1, \ldots, T_k$ are independent with survival functions, respectively, $e^{-zH_j(t)}$ for given parametric functions $H_j(t)$. In this case the joint survival function becomes

$$\bar{K}(t_1, \ldots, t_k) = \int_0^\infty \exp\left\{-z\sum_{j=1}^k H_j(t)\right\} dV(z)$$

where $V(z)$ is the distribution function of Z. As an example, letting V correspond to a gamma-distribution with expected value 1 and variance δ, we get

$$\bar{K}(t_1, \ldots, t_k) = \left(1 + \delta \sum_{j=1}^k H_j(t)\right)^{-1/\delta}$$

Note that when $\delta \to 0$ this tends to $\exp\{-z\sum_{j=1}^k H_j(t)\}$ which is the model obtained under independent risks. Also, if the $H_j(t)$ are of the Weibull form $(t/\theta_j)^{\alpha_j}$, we arrive at the multivariate Burr distribution (see Crowder [11, p. 43]).

Preventive maintenance modelling

In subsection 1.2 we considered the case with $k = 2$ where T_1, T_2 are, respectively, the time of critical failure of a component and the potential time of a PM. Cooke [7], [9] introduced the notion of random signs censoring which is tailored for such cases. In our notation the PM-time T_2 is called a random signs censoring of the failure time T_1 if the event $\{T_2 < T_1\}$ is stochastically independent of T_1. Thus, random signs censoring means that the event that the failure of the component is preceded by PM, is not influenced by the time T_1 at which the component fails or would have failed without PM. The idea is that the component emits some kind of signal before failure, and that this signal is discovered with a probability which does not depend on the age of the component. Moreover, random signs censoring implies identifiability of the distribution of T_1.

Lindqvist, Støve and Langseth [21] suggested a model called the repair alert model for describing the joint behavior of failure times T_1 and PM-times T_2. This model is a special case of random signs censoring obtained by introducing a repair alert function which describes the "alertness" of the maintenance crew as a function of time.

5 Statistical inference

Consider the case where we have data of the form (T, C) for n independent observation units. In practice such data are often right censored by some source independent of the k given risks. For example, this could be a censoring due to a time limit of the experiment (type I censoring). Then if the ith observation is non-censored, we observe both the lifetime t_i and the cause c_i. On the other hand, if the ith observation is right censored at time t_i, then we do not observe c_i and all we know is that $T > t_i$. If we let $\delta_i = 0$ if the ith observation is censored and $\delta_i = 1$ otherwise, we get the likelihood function

$$L = \prod_{i=1}^{n} f_{c_i}(t_i)^{\delta_i} \bar{F}(t_i)^{1-\delta_i}$$

Using equation (1) we can write this as

$$L = \prod_{i=1}^{n} \lambda_{C_i}(t_i)^{\delta_i} \bar{F}(t_i)$$

Further, writing $\lambda_{C_i}(t_i)^{\delta_i} = \prod_{j=1}^{k} \lambda_j(t_i)^{\delta_{ij}}$, where $\delta_{ij} = I(C_i = j)$ when $\delta_i = 1$ and $\delta_{ij} = 0$ when $\delta_i = 0$, and then using equation (4), we arrive at (see Lawless [19, Ch. 9.1])

$$L = \prod_{j=1}^{k} \left[\prod_{i=1}^{n} g_j^*(t_i)^{\delta_{ij}} G_j^*(t_i)^{1-\delta_{ij}} \right] \equiv \prod_{j=1}^{k} L_j \tag{12}$$

where $g_j^*(t) = \lambda_j(t)\bar{G}_j^*(t)$ is the density function corresponding to $\bar{G}_j^*(t)$. Thus we can write L as a product of the likelihoods L_j, where L_j has the same form as the likelihood function of a censored sample associated with the jth failure cause where all observations where C is not observed to equal j are treated as censorings. This fact leads to a simplification of parametric estimation problems if the sub-distributions are modelled by parameters which are separate for each failure cause (for example the model (11)).

Equation (12) has, furthermore, important implications for non-parametric estimation. In fact, the likelihood L_j is of the same form as the one leading to the usual Kaplan-Meier estimator (see [2]) of a survival function under independent right censoring. This means in particular that the "artificial" survival functions $\bar{G}_j^*(t)$ and hence $\Lambda_j(t) = -\log \bar{G}_j^*(t)$, can be estimated non-parametrically by the Kaplan-Meier estimator. Alternatively, one may from the same reasoning estimate the Λ_j for each j using the Nelson-Aalen estimator ([2]). Denoting the resulting estimator by $\hat{\Lambda}_j(t)$, it follows from equation (3) that one may estimate the F_j non-parametrically by

$$\hat{F}_j(t) = \int_0^t \hat{\bar{F}}(u) \mathrm{d}\hat{\Lambda}_j(u), \; j = 1, ..., k, . \tag{13}$$

where $\hat{\bar{F}}(t)$ is the ordinary Kaplan-Meier estimate of the survival function $\bar{F}(t)$ of T. Note that this estimator is based on the censored data obtaining by collapsing the k failure causes into one single cause.

It should be noted that in the case of independent risks, then the marginal survival distributions are always consistently estimated by the Kaplan-Meier estimator for each of the risks. It is interesting to note that Zheng and Klein [31], in the case $k = 2$ and with a known copula describing the dependence of T_1 and T_2, derive non-parametric estimators of the marginal survival distributions which specialize to the Kaplan-Meier estimator in the case of independent risks.

5.1 Beyond classical competing risks

Markov process models

Aalen [1] modelled the classical competing risks problem as a continuous-time Markov process with one working state ("0") and k absorbing failure states corresponding to the k risks. More generally one may consider Markov processes with more complex state spaces and again k absorbing states corresponding to the k failure types. This leads to the consideration of multivariate phase type distributions, see Assaf et al. [3]. Hokstad and Frøvig [16] consider failure models for periodically tested items where the failures may be either degraded or critical. Whitmore [29] discusses the use of first-passage-time distributions connected with multidimensional Brownian motion as models for competing risks.

Competing risks in repairable systems

Consider a system where failures are classified in k different types. Until now we have considered the case of non-repairable units. Thus assume that the unit is repaired after failure, then put into operation again, and that the process continues in this way. At each failure the type of event is recorded. Assume also that time durations of repair and maintenance can be disregarded, so that the system is always restarted immediately after failure or a maintenance action This leads to the observation of a marked point process $(S_1, C_1), (S_2, C_2), \ldots$ with successive failure times $0 < S_1 < S_2 < \cdots$ and marks C_i in $\{1, 2, \ldots, k\}$. The properties of such a process depends on the repair strategy. For example, a perfect repair (renewal) of the system at failures leads to independent and identically distributed pairs (T_i, C_i) where $T_i = S_i - S_{i-1}$ $(i = 1, 2, \ldots)$ where $S_0 \equiv 0$. Doyen and Gaudoin [13] present a point process approach for modelling of imperfect repair in competing risks situations between failure and PM. Bedford and Lindqvist [5] considered a series system of k repairable components where only the failing component is repaired at failures. A general setup for this kind of preocesses is suggested by Lindqvist [20].

6 Discussion

It has been argued above that the latent failure time approach is particularly useful in reliability applications. Here the latent times may be, for example, times to failure of a certain system component or time to PM. However, the risks will usually be dependent so that the unidentifiability problem occurs. It should be noted that, although additional assumptions on the dependence may lead to identifiability, these assumptions are non-testable from data of the form (T, C). Prentice et al. [27] therefore rejected the latent failure time approach and instead advocated the use of the observable sub-hazard functions in analyses of competing risks. One may argue against this, however, that in practical applications it may often make good sense to use ones information and prior beliefs in order to model the underlying probability mechanisms beyond what is actually observable.

On the other hand, an erroneous assumption of independent risks may lead to seriously misleading conclusions. As noted by Cooke [8], such assumptions are often done in the assessments of competing failure rates in reliability databases. One then assumes that for each of the k failure causes, failures occur as independent Poisson processes. However, this implies that the rate of occurrence of each of the competing risks would be unaffected by removing the others. For competing risks corresponding to critical failure and PM this means that the rate of occurrence of critical failures would be unaffected by stopping preventive maintenance activity, an assumption which is completely unreasonable. The appropriate alternative method would be to invoke a more careful modelling by competing risks, using if possible all available information. More sophisticated methods for repairable systems as briefly mentioned above may be needed here.

References

[1] Aalen O O. Nonparametric inference in connection with multiple decrement models. Scandinavian Journal of Statistics 1976 3: 15–27.

[2] Andersen P, Borgan O, Gill R, Keiding N. Statistical Models Based on Counting Processes; Springer: New York, 1992.

[3] Assaf D, Langberg N A, Savits T H, Shaked, M. Multivariate phase-type distributions. Operations Research 1984 32: 688–702.

[4] Bedford T, Cooke R M. Probabilistic Risk Analysis: Foundations and Methods; Cambridge University Press: Cambridge, 2001.

[5] Bedford T, Lindqvist B H. The identifiability problem for repairable systems subject to competing risks. Advances in Applied Probability 2004 36: 774–790.

[6] Bedford T, Meilijson I. A characterization of marginal distributions of (possibly dependent) lifetime variables which right censor each other. Annals of Statistics 1997 25: 1622–1645.

[7] Cooke R M. The total time on test statistics and age-dependent censoring. Statistics and Probability Letters 1993 18: 307–312.

[8] Cooke R M. The design of reliability databases, Part I. Reliability Engineering and System Safety 1996 51: 137–146.

[9] Cooke R M. The design of reliability databases, Part II. Reliability Engineering and System Safety 1996 51: 209–223.

[10] Cox D R. The analysis of exponentially distributed lifetimes with two types of failure. Journal of Royal Statistical Society Series B 1959 21: 411–421.

[11] Crowder M J. Classical competing risks; Chapman & Hall/CRC: Boca Raton, 2001.

[12] David H A, Moeschberger M L. The Theory of Competing Risks; Griffin: London, 1978.

[13] Doyen L, Gaudoin O. Imperfect maintenance in a generalized competing risks framework. To appear in Journal of Applied Probability 2006 43.

[14] Gail M. A review and critique of some models used in competing risks analysis. Biometrics 1975 31: 209–222.

[15] Gumbel E J. Bivariate exponential distributions. Journal of American Statistical Association 1960 55: 698–707.

[16] Hokstad P, Frøvig A T. The modelling of degraded and critical failures for components with dormant failures. Reliability Engineering and System Safety 1996 51: 189–199.

[17] Kalbfleisch JD, Prentice R L. The Statistical Analysis of Failure Time Data; John Wiley & Sons: New York, 1980.

[18] King J R. Probability Charts for Decision Making; Industrial Press: New York, 1971.

[19] Lawless J F. Statistical models and methods for lifetime data, 2nd ed. Wiley-Interscience: Hoboken NJ, 2003.

[20] Lindqvist B H. On the statistical modelling and analysis of repairable systems. Accepted for publication in Statistical Science, 2006.

[21] Lindqvist B H, Støve B, Langseth H. Modelling of dependence between critical failure and preventive maintenance: The repair alert model. Journal of Statistical Planning and Inference 2006 136: 1701–1717.

[22] Mendenhall W, Hader R J. Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. Biometrika 1958 45: 504–520.

[23] Moeschberger M L, Klein J P. Statistical methods for dependent competing risks. Lifetime Data Analysis 1995 1: 195–204.

[24] Nelsen R B. An Introduction to Copulas; Springer Verlag: New York, 1999.

[25] Nelson W. Applied Life Data Analysis; John Wiley & Sons: New York, 1982.

[26] Peterson A V. Bounds for a joint distribution function with fixed sub-distribution functions: application to competing risks. Proceedings of National Academy of Sciences USA 1976 73: 11–13.

[27] Prentice R L, Kalbfleisch J D, Peterson A V, Flournoy N, Farewell V T, Breslow N E. The analysis of failure times in the presence of competing risks. Biometrics 1978 34: 541–554.

[28] Tsiatis A. A nonidentifiability aspect of the problem of competing risks. Proceedings of National Academy of Sciences USA 1975 72: 20–22.

[29] Whitmore G A. First-passage time models for duration data: regression structures and competing risks. The Statistician 1986 35: 207–219.

[30] Williams J S, Lagakos S W. Models for censored survival analysis: constant- sum and variable-sum models. Biometrika 1977 64: 215–224.

[31] Zheng M, Klein J P. Estimates of marginal survival for dependent competing risks based on an assumed copula. Biometrika 1995 82: 127–138.