Biometrical Journal **40** (1998) 8, 963–978

An Asymptotically Optimal Adaptive Selection Procedure in the Proportional Hazards Model with Conditionally Independent Censoring*

ANDREAS WIENKE

Max Planck Institute for Demographic Research
Rostock
Germany

Summary

Assume k independent populations are given which are distributed according to $R_{\vartheta_1}, \ldots, R_{\vartheta_k}$ ($\vartheta_i \in \Theta \subseteq \boldsymbol{R}$). Taking samples of size n the population with the smallest ϑ-value is to be selected. Using the framework of Le Cam's decision theory (LE CAM, 1986; STRASSER, 1985) under mild regularity assumptions, an asymptotically optimal selection procedure is derived for the sequence of localized models. In the proportional hazards model with conditionally independent censoring, an asymptotically optimal adaptive selection procedure is constructed by substituting the unknown nuisance parameter by a kernel estimator.

Key words: Selection procedure; Conditionally independent censoring; Proportional hazards model.

1. Introduction

Suppose that $\pi_1, \ldots, \pi_k$ are independent populations with distributions $R_{\vartheta_1}, \ldots, R_{\vartheta_k}$ ($\vartheta_i \in \Theta \subseteq \boldsymbol{R}$). The population with the smallest ϑ-value is called the best population. As in the theory of hypothesis testing, only under special conditions one can find finite optimal selection rules based on samples from the populations. Therefore, we are looking for asymptotically optimal solutions to this problem. To achieve this we apply the results of asymptotic decision theory.

If there is exactly one smallest ϑ-value then for every reasonable sequence of selection procedures $\{q_n\}$ the probability of correct selection $P(CS, q_n)$ tends to one if the sample size n (assumed to be equal for all populations) goes to infinity. This makes a direct comparison between two sequences $\{q_n^1\}$ and $\{q_n^2\}$ of selection rules impossible. Similar to the theory of hypothesis testing we consider localized models $R_{\vartheta_0+\frac{h_1}{\sqrt{n}}}, \ldots, R_{\vartheta_0+\frac{h_k}{\sqrt{n}}}$. We are now interested in selecting the population with the smallest h-value. In this case the probability of correct selection con-

* presented at the 44[th] Biometrical Colloquium of the German Region of the International Biometric Society, Mainz, 1998

verges to a value between zero and one and one can compare the efficiency of $\{q_n^1\}$ and $\{q_n^2\}$ by the limits $\lim\limits_{n\to\infty} P(CS, q_n^i)$.

LIESE (1996) used the general Hajek-Le Cam-bound of asymptotic decision theory to establish an upper bound for the minimum probability of correct selection for a general class of models following the indifference zone approach of BECHHOFER (1954). A sequence of selection rules is called asymptotically optimal if it attains this upper bound. LIESE (1996) derived asymptotically optimal selection procedures in the location model. We assume a proportional hazards model which includes a infinite dimensional nuisance parameter. Furthermore the lifetimes and censoring times are assumed to be conditionally independent given a finite dimensional covariate. For this model we construct a selection procedure without any prior information on the nuisance parameter and show that this procedure has the same asymptotic efficiency as the best selection rule with known nuisance parameter.

2. Selection Procedures and Decision Theory

From each of the populations $\pi_1, \dots, \pi_k$ we take a sample of n independent observations $V_{i1}, \dots, V_{in}$ $(i = 1, \dots, k)$. To be more general we assume first of all that the V_{ij} are random variables on a suitable measurable space $(\Omega, \mathcal{F})$. In the special case of real variables $(\Omega, \mathcal{F}) = (\boldsymbol{R}, \mathcal{B})$ holds, where $\mathcal{B}$ denotes the σ-algebra of borel sets on $\boldsymbol{R}$. Let $\{R_{\vartheta_i},\ \vartheta_i \in \Theta \subseteq \boldsymbol{R}\}$ be the distribution of the measurements of population π_i. Let $\Theta^k = \Theta \times \ldots \times \Theta$ denote the k-dimensional parameter space of the parameter vector $\theta = (\vartheta_1, \dots, \vartheta_k)$ and $R_\vartheta^n = R_\vartheta \times \ldots \times R_\vartheta$ the n-times product measure of R_ϑ. To simplify the notation we set $P_\theta^n = R_{\vartheta_1}^n \times \ldots \times R_{\vartheta_k}^n$. If we treat the problem of selecting the best population as a decision problem the corresponding decision space is $\boldsymbol{D} = \{1, \dots, k\}$.

Definition: A selection rule q is a discrete probability distribution

$$q(v) = (q_1(v), \dots, q_k(v)), \qquad q_i(v) \geq 0, \qquad \sum_{i=1}^k q_i(v) = 1, \qquad v \in \Omega^{nk},$$

where q_i denotes the probability of selecting population π_i.

Let $\vartheta_{[1]} \leq \ldots \leq \vartheta_{[k]}$ be the ordered values of the parameters $\vartheta_1, \dots, \vartheta_k$. The selection of the population with the smallest parameter value $\vartheta_{[1]}$ is called correct selection (CS). Following Bechhofer's indifference zone approach we introduce the preference zone for fixed $\gamma > 0$ as $\Theta_\gamma = \{\theta = (\vartheta_1, \dots, \vartheta_k) : \theta \in \Theta^k, \vartheta_{[1]} \leq \vartheta_{[2]} - \gamma\}$. Only in special cases (for example in exponential families) do uniformly best (permutation invariant) selection rules exist. Therefore we focus on selection procedures q^*, which guarantee high probability of correct selection in the worst case of parameter configuration (least favorable configuration), so-called minimax solutions of the form $\inf\limits_{\theta\in\Theta_\gamma} P_\theta^n(CS, q^*) \geq \inf\limits_{\theta\in\Theta_\gamma} P_\theta^n(CS, q)$. To study the minimax risk we systematically employ results from asymptotic decision theory. We have to impose some regularity conditions on the distribution family $\{R_\vartheta,\ \vartheta \in \Theta\}$, to which the distributions $R_{\vartheta_1}, \ldots, R_{\vartheta_k}$ of the populations $\pi_1, \ldots, \pi_k$ belong. Furthermore, let $I(\vartheta)$ be the Fisher information of $\{R_\vartheta,\ \vartheta \in \Theta\}$ and suppose $\dot{\Theta}$ is not empty, where $\dot{\Theta}$ is the interior of Θ.

Assumption (A1): The distribution family $\{R_\vartheta,\ \vartheta \in \Theta\}$ is $L_2(\vartheta_0)$-differentiable with derivation $\dot{l}_{\vartheta_0}$ and Fisher information $I(\vartheta_0) > 0 \quad \forall \vartheta_0 \in \dot{\Theta}$.

For the notion of L_2-differentiability we refer to WITTING (1985). Similar to the asymptotic theory of statistical tests we localize our experiment. To be more precise we fix $\theta_0 = (\vartheta_0, \ldots, \vartheta_0) \in \dot{\Theta} \times \ldots \times \dot{\Theta}$. Note that for each population the corresponding model is localized at the same ϑ_0. Set

$$H_n = \left\{ \boldsymbol{h} : \boldsymbol{h} = (h_1, \ldots, h_k) \in \boldsymbol{R}^k,\ \left(\vartheta_0 + \frac{h_1}{\sqrt{n}}, \ldots, \vartheta_0 + \frac{h_k}{\sqrt{n}} \right) \in \dot{\Theta} \times \ldots \times \dot{\Theta} \right\},$$

and $P_{n,\boldsymbol{h}} = R^n_{\vartheta_0 + \frac{h_1}{\sqrt{n}}} \times \ldots \times R^n_{\vartheta_0 + \frac{h_k}{\sqrt{n}}}$ $(\boldsymbol{h} \in H_n)$, and introduce the statistical experiment for our selection problem as

$$\mathcal{E}_n = (\Omega^{nk}, \mathcal{F}^{nk}, P_{n,\boldsymbol{h}}, \boldsymbol{h} \in H_n)\,. \tag{1}$$

A first step is to derive an upper bound for $\inf\limits_{\boldsymbol{h} \in H_n} P_{n,\boldsymbol{h}}(CS, q_n)$ for any sequence of selection rules. This can be done with the help of the Hajek-Le Cam-bound of asymptotic decision theory. Introduce $H_{n,\eta} = \{\boldsymbol{h} : \boldsymbol{h} \in H_n,\ h_{[1]} \le h_{[2]} - \eta\}$ and denote by K_c the ball around the origin with radius $c > \eta$. It holds (LIESE (1996), Theorem 1) that

$$\limsup_{n\to\infty} \inf_{\boldsymbol{h} \in H_{n,\eta} \cap K_c} P_{n,\boldsymbol{h}}(CS, q_n) \le \int \Phi^{k-1}(x + \eta\sqrt{I(\vartheta_0)})\, \Phi(\mathrm{d}x) \tag{2}$$

with Φ as the distribution function of the standard normal distribution. A next step is to construct a sequence of selection rules which attains the upper bound and is optimal in this sense. Using the notation $V = (V_1, \ldots, V_k)$ and $V_i = (V_{i1}, \ldots, V_{in})$ the central sequence of $\mathcal{E}_n$ from (1) is a k-dimensional vector $Z_n = (Z_{1,n}, \ldots, Z_{k,n})$ with $Z_{i,n}(V) = \dfrac{1}{\sqrt{n}\, I(\vartheta_0)} \sum\limits_{j=1}^{n} \dot{l}_{\vartheta_0}(V_{ij})$ $(i = 1, \ldots, k)$.

Let $\#A$ denote the cardinality of a set A. This leads to the following theorem (see LIESE (1996), Proposition 1):

Theorem 1: *Let assumption* **(A1)** *be fulfilled and define* $A(Z_n) \subseteq \{1, \ldots, k\}$ *by*

$$A(Z_n) := \{i : Z_{i,n} = \min_{1 \le j \le k} Z_{j,n}\}\,. \tag{3}$$

Introduce $q_n(Z_n) = \big(q_{1,n}(Z_n), \ldots, q_{k,n}(Z_n)\big)$ *with*

$$q_{i,n}(Z_n) = \begin{cases} \dfrac{1}{\#A(Z_n)} & \textit{if} \quad i \in A(Z_n) \\ 0 & \textit{if} \quad i \notin A(Z_n) \end{cases} \tag{4}$$

the discrete uniform distribution on $A(Z_n)$, i.e. it is the rule which selects the population according to the smallest component in the central sequence and break ties randomly. Then $\{q_n\}$ is an asymptotically optimal selection rule for the selection problem $\mathcal{E}_n$ in (1) in the sense that it attains the upper bound in (2).

The above theorem is the key to deriving asymptotically optimal selection procedures. For concrete models it is necessary to calculate the central sequence explicitly. This is the content of the following section.

3. The Proportional Hazards Model

Let $X_{i1}, \ldots, X_{in}$ $(i = 1, \ldots, k)$ be lifetimes (nonnegative continuous random variables) from k independent populations $\pi_1, \ldots, \pi_k$. For example the data may represent the lifetimes of k groups of patients with different treatments in a clinical trial. Often the experimenter is interested in ascertaining which population is associated with the longest lifetimes. The parameters $\vartheta_1, \ldots, \vartheta_k$ characterize the quality of the populations $\pi_1, \ldots, \pi_k$. In a stochastically decreasing model like the proportional hazards model the population with the smallest ϑ-value generates the largest random variables, which is why this population is called the best.

In practice is it sometimes impossible to measure lifetimes directly due to the occurrence of censoring events. The present model includes such incomplete data in case of conditionally independent censoring.

Let $Y_{i1}, \ldots, Y_{in}$ $(i = 1, \ldots, k)$ be the censoring times. The X_{ij} are not directly observable, rather, one is able to observe only (T_{ij}, Δ_{ij}), where $T_{ij} = \min\{X_{ij}, Y_{ij}\}$ and $\Delta_{ij} = 1(X_{ij} \leq Y_{ij})$ is a binary random variable. It turns out that in the present model asymptotically optimal selection procedures depend on the (unknown) survival function of the lifetimes. The problem of estimating the survival function in the presence of random right censoring has been extensively studied. Most research has centered on the independent censoring model, in which the censoring times Y_{ij} are stochastically independent of the lifetimes X_{ij}. Under this model the observable data (T_{ij}, Δ_{ij}) provide sufficient information to uniquely determine the marginal distribution of X_{ij} and the Kaplan-Meier-estimate (KAPLAN and MEIER, 1958) (KME) is the appropriate estimate of the survival function.

But the assumption of independence between lifetimes and censoring times is not always true. LAGAKOS (1979) mentioned the following example of a situation, in which the independence assumption is of questionable validity: a clinical trial in which those patients experiencing a specific critical event such as metastatic spread of disease are, by design, removed from study and no longer followed for survival time.

If the only observations available are the pairs (T_{ij}, Δ_{ij}), the independence assumption is completely untestable. It has been shown by TSIATIS (1975) "that there always exist independent censoring models consistent with any probability distribution for the observable pair (T, Δ)" (Lagakos). For an excellent review of the field of dependent censoring see MOESCHENBERGER and KLEIN (1995).

For these reasons we introduce a model that was motivated by CHENG (1989) and ARNOLD and KIM (1992). This model includes a finite dimensional covariate C such as age, blood pressure, body mass index or alcohol intake in addition to the (censored) lifetimes of the patients. The observed sample information is now $V_{ij} = (T_{ij}, \Delta_{ij}, C_{ij})$ with V_{ij} from section 2.

It is necessary to specify the type of dependence between lifetimes and censoring times. We assume that X and Y are conditionally independent given $C = c$. Let P_C be the distribution of the covariate C, $P_\vartheta(\cdot \mid c)$ and $Q(\cdot \mid c)$ denote the conditional distributions of lifetimes and censoring times given $C = c$ and $F_\vartheta(x \mid c)$ and $G(y \mid c)$ be the cumulative distribution functions (c.d.f.) of $P_\vartheta(\cdot \mid c)$ and $Q(\cdot \mid c)$, respectively. We suppose a proportional hazards model of the form

$$F_{\vartheta_i}(x \mid c) = 1 - (1 - F^o(x \mid c))^{\vartheta_i} \qquad (\vartheta_i \in \Theta = (0, \infty)) \tag{5}$$

for the lifetimes of the populations $\pi_1, \ldots, \pi_k$ with an unknown basic c.d.f. $F^o(x \mid c)$ (w.r.t. x). $P^o(\cdot \mid c)$ stands for the distribution associated with $F^o(x \mid c)$. Furthermore, it is necessary to make the following assumptions:

Assumption (A2): $P^o(\cdot \mid c)$, $Q(\cdot \mid c)$ and $P_C(\cdot)$ are absolutely continuous distributions with densities $f^o(x \mid c)$, $g(x \mid c)$ and $p_C(c)$. Additionally it holds that $P(X \leq Y) > 0$.

Note that $P(X \leq Y) > 0$ means censoring does not occur with probability one. In view of **(A2)** there exists a conditional density function $f_\vartheta(x \mid c)$ w.r.t. $F_\vartheta(x \mid c)$.

Lemma 1: *Let the assumption* **(A2)** *be fulfilled. Then the three-dimensional distribution family* $\{P^*_\vartheta, \vartheta \in (0, \infty)\}$ *of* (X, Y, C) *is* $L_2(\vartheta_0)$*-differentiable* $\forall \vartheta_0 \in (0, \infty)$ *with* L_2*-derivation* $\dot{L}_{\vartheta_0}(x, c) = \frac{1}{\vartheta_0}\big(1 + \ln(1 - F_{\vartheta_0}(x \mid c))\big)$ *and Fisher information* $I(\vartheta_0) = 1/\vartheta_0^2$.

Proof: Because of assumption **(A2)** the density of the random vector (X, Y, C) in the proportional hazards model is of the form $h_\vartheta(x, y, c) = f_\vartheta(x \mid c)\, g(y \mid c)\, p_C(c)$. In view of WITTING (1985), Theorem 1.194 it is sufficient to show the continuity of the Fisher information w.r.t. the parameter ϑ. The derivation of the density w.r.t. ϑ is of the form $\frac{h_\vartheta(x, y, c)}{d\vartheta} = \dot{h}_\vartheta(x, y, c) = \dot{f}_\vartheta(x \mid c) \times g(y \mid c)\, p_C(c)$ and consequently $\frac{\dot{h}_{\vartheta_0}}{h_{\vartheta_0}}(x, y, c) = \frac{\dot{f}_{\vartheta_0}}{f_{\vartheta_0}}(x \mid c)$. Furthermore, $f_\vartheta(x \mid c) = \vartheta f^o(x \mid c)\,(1 - F^o(x \mid c))^{\vartheta - 1}$ and

$$\begin{aligned}\dot{f}_\vartheta(x \mid c) = {} & f^o(x \mid c)\,(1 - F^o(x \mid c))^{\vartheta - 1} \\ & + \vartheta(1 - F^o(x \mid c))^{\vartheta - 1} \ln(1 - F^o(x \mid c))\,.\end{aligned}$$

Hence

$$\dot{L}_{\vartheta_0}(x,c) = \frac{\dot{f}_{\vartheta_0}}{f_{\vartheta_0}}\,(x \mid c) = \frac{1}{\vartheta_0} + \ln\left(1 - F^o(x \mid c)\right)$$
$$= \frac{1}{\vartheta_0}\left(1 + \ln\left(1 - F_{\vartheta_0}(x \mid c)\right)\right)$$

and

$$I(\vartheta_0) = \frac{1}{\vartheta_0^2}\int\int \left(1 + \ln\left(1 - F_{\vartheta_0}(x \mid c)\right)\right)^2 P_{\vartheta_0}(\mathrm{d}x \mid c)\, P_C(\mathrm{d}c) = \frac{1}{\vartheta_0^2}$$

which completes the proof. ■

The following lemma is a modification of Lemma 2 from JANSSEN (1989). It gives the L_2-derivation of the distribution family $\{R_\vartheta, \vartheta \in \Theta\}$, to which the observable data $V = (T, \Delta, C)$ belong.

Lemma 2: *Let assumption* **(A2)** *be fulfilled. Then* $\{R_\vartheta,\ \vartheta \in \Theta\}$ *is* L_2*-differentiable with derivation*

$$\dot{l}_{\vartheta_0}(t,\delta,c) = \delta \dot{L}_{\vartheta_0}(t,c) + (1-\delta)\int\limits_{(t,\infty)} \dot{L}_{\vartheta_0}(x,c)\, P_{\vartheta_0}(\mathrm{d}x \mid c)\big/\left(1 - F_{\vartheta_0}(t \mid c)\right).$$

Note that the regularity assumptions are made on the distribution of the non-observable random variable (X, Y, C). To apply Theorem 1 it is necessary to check these assumptions w.r.t. the distribution of the censored observation (T, Δ, C). This is the content of the next lemma. The following table illustrates the notation used for observable and non-observable data:

Table 1
Notation for observable and non-observable data

	non-observable data	observable data
random variable	(X, Y, C)	(T, Δ, C)
distribution	P_ϑ^*	R_ϑ
L_2-derivation	$\dot{L}_\vartheta$	$\dot{l}_\vartheta$

Lemma 3: *Let assumption* **(A2)** *be fulfilled. Then for the proportional hazards model (5) the family* $\{R_\vartheta,\ \vartheta \in (0,\infty)\}$ *is* $L_2(\vartheta_0)$*-differentiable with derivation*

$$\dot{l}_{\vartheta_0}(t,\delta,c) = \frac{1}{\vartheta_0}\left(\delta + \ln\left(1 - F_{\vartheta_0}(t \mid c)\right)\right) \tag{6}$$

and it holds that

$$I(\vartheta_0) = \frac{1}{\vartheta_0^2}\int\int \left(1 - G(x \mid c)\right) P_{\vartheta_0}(\mathrm{d}x \mid c)\, P_C(\mathrm{d}c) > 0 \qquad \forall \vartheta_0 \in (0,\infty)\,.$$

Proof: Lemma 1 and Lemma 2 imply for the L_2-derivation of the proportional hazards model with conditional independent censoring

$$\dot{l}_{\vartheta_0}(t,\delta,c) = \delta\left(\frac{1}{\vartheta_0} + \ln\left(1 - F^o(t \mid c)\right)\right) + (1-\delta)\,\frac{\displaystyle\int_{(t,\infty)} \left(\frac{1}{\vartheta_0} + \ln\left(1 - F^o(x \mid c)\right)\right) P_{\vartheta_0}(\mathrm{d}x \mid c)}{1 - F_{\vartheta_0}(t \mid c)}$$

$$= \frac{1}{\vartheta_0}\left(\delta + \ln\left(1 - F_{\vartheta_0}(t \mid c)\right)\right). \tag{7}$$

For the Fisher information the following holds in view of the conditional independence

$$\begin{aligned} I(\vartheta_0) &= \iiint \dot{l}^2_{\vartheta_0}(t,\delta,c)\, R_{\vartheta_0}(\mathrm{d}t,\mathrm{d}\delta,\mathrm{d}c) \\ &= \frac{1}{\vartheta_0^2}\iint \ln^2\left(1 - F_{\vartheta_0}(t \mid c)\right)\left(1 - F_{\vartheta_0}(t \mid c)\right) Q(\mathrm{d}t \mid c)\, P_C(\mathrm{d}c) \\ &\quad + \frac{1}{\vartheta_0^2}\iint \left(1 + \ln\left(1 - F_{\vartheta_0}(t \mid c)\right)\right)^2 \left(1 - G(t \mid c)\right) P_{\vartheta_0}(\mathrm{d}t \mid c)\, P_C(\mathrm{d}c). \end{aligned}$$

Integrating-by-parts yields

$$\begin{aligned} &\int \left(1 + \ln\left(1 - F_{\vartheta_0}(t \mid c)\right)\right)^2 \left(1 - G(t \mid c)\right) P_{\vartheta_0}(\mathrm{d}t \mid c) \\ &= \int \left(1 - G(t \mid c)\right) P_{\vartheta_0}(\mathrm{d}t \mid c) \\ &\quad - \int \ln^2\left(1 - F_{\vartheta_0}(t \mid c)\right)\left(1 - F_{\vartheta_0}(t \mid c)\right) Q(\mathrm{d}t \mid c). \qquad \blacksquare \end{aligned}$$

Using the notation $T_i = (T_{i1}, \ldots, T_{in})$, $\Delta_i = (\Delta_{i1}, \ldots, \Delta_{in})$, $C_i = (C_{i1}, \ldots, C_{in})$ Lemma 2 and Theorem 1 lead to the following theorem, which gives the concrete form of optimal selection procedures in the model under consideration:

Theorem 2: *Let assumption* **(A2)** *be fulfilled and introduce the k-dimensional statistic* $S_n = (S_{1,n}(T_1, \Delta_1, C_1), \ldots, S_{k,n}(T_k, \Delta_k, C_k))$ *by*

$$S_{i,n}(T_i, \Delta_i, C_i) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\Delta_{ij} + \ln\left(1 - F_{\vartheta_0}(T_{ij} \mid C_{ij})\right)\right) \tag{8}$$

for $i = 1, \ldots, k$. *Then the selection rule* $q_n(S_n) = (q_{1,n}(S_n), \ldots, q_{k,n}(S_n))$ *with*

$$q_{i,n}(S_n) = \begin{cases} \dfrac{1}{\#A(S_n)} & \text{if} \quad i \in A(S_n) \\ 0 & \text{if} \quad i \notin A(S_n) \end{cases} \tag{9}$$

is asymptotically optimal in the proportional hazards model (5) with conditionally independent censoring in the sense that it attains the upper bound in (2).

Note that the optimal selection procedure does not depend on the censoring distribution. This changes if the censoring distribution is not equal over all populations.

4. Adaptive Selection

An asymptotically optimal selection rule was given in Theorem 2. Unfortunately, the statistics $S_{i,n}$ in (8) include the survival function $1 - F_{\vartheta_0}(t \mid c) = (1 - F^o(t \mid c))^{\vartheta_0}$. The essential point now is to replace this unknown function with a reasonable estimate. Because of the dependence between lifetimes and censoring times the KME is not a consistent estimate in this model. It turns out that it is easier to estimate the cumulative hazard function instead of $1 - F_{\vartheta_0}(t \mid c)$. Let

$$\Lambda(t \mid c) = -\ln\left(1 - F_{\vartheta_0}(t \mid c)\right) = \int_0^t \frac{P_{\vartheta_0}(\mathrm{d}x \mid c)}{1 - F_{\vartheta_0}(x \mid c)}$$

be the conditional cumulative hazard function (c.h.f.) of $F_{\vartheta_0}(t \mid c)$ given $C = c$. Under the conditional independence assumption the upper expression is identifiable (BERAN, 1981). Using the notation $A(t, c) := p_C(c) \int_0^t (1 - G(x \mid c))\, P_{\vartheta_0}(\mathrm{d}x \mid c)$ and $B(t, c) := p_C(c)\,(1 - G(t \mid c))\,(1 - F_{\vartheta_0}(t \mid c))$ the conditional c.h.f. can be written in the form

$$\Lambda(t \mid c) = \int_0^t \frac{p_C(c)\,(1 - G(x \mid c))\, P_{\vartheta_0}(\mathrm{d}x \mid c)}{p_C(c)\,(1 - G(x \mid c))\,(1 - F_{\vartheta_0}(x \mid c))} = \int_0^t \frac{A(\mathrm{d}x, c)}{B(x, c)}. \tag{10}$$

Let $K_{b_N}(x, y) = \dfrac{1}{b_N}\, K\left(\dfrac{x - y}{b_N}\right)$ be a kernel with bandwidth b_N. Note that the functions A and B can be written in the form $A(t, c) = p_C(c)\, P(T < t, \Delta = 1 \mid c)$ and $B(t, c) = p_C(c)\, P(T \geq t \mid c)$. Consequently, their natural estimates are

$$A_N(t, c) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} 1(T_{ij} < t, \Delta_{ij} = 1)\, K_{b_N}(C_{ij}, c) \tag{11}$$

and

$$B_N(t, c) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} 1(T_{ij} \geq t)\, K_{b_N}(C_{ij}, c)\,. \tag{12}$$

$N = kn$ denotes the whole sample size from all k populations. Because of (10)

$$\hat{\Lambda}_N(t \mid c) := \int_0^t \frac{A_N(\mathrm{d}x, c)}{\varepsilon_N + B_N(x, c)} \tag{13}$$

is a reasonable estimate for the conditional c.h.f. Here ε_N denotes a monotone decreasing sequence of positive real numbers converging to zero. ε_N is only used to keep the denominator away from zero. Furthermore we introduce the sequence of statistics $\Lambda_N(t \mid c) = \int\limits_0^t \frac{A(\mathrm{d}x, c)}{\varepsilon_N + B(x, c)}$ with ε_N from (13).

Theorem 3: *Let assumption* **(A2)** *be fulfilled, let* $\varepsilon_N > 0$ *be a monotone sequence with* $\lim\limits_{N \to \infty} \varepsilon_N = 0$ *and introduce* $\tilde{S}_n := (\tilde{S}_{1,n}(T_1, \Delta_1, C_1), \ldots, \tilde{S}_{k,n}(T_k, \Delta_k, C_k))$ *by*

$$\tilde{S}_{i,n}(T_i, \Delta_i, C_i) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\Delta_{ij} - \Lambda_N(T_{ij} \mid C_{ij})) \tag{14}$$

for $i = 1, \ldots, k$. *Then the selection rule* $\tilde{q}_n = (\tilde{q}_{1,n}, \ldots, \tilde{q}_{k,n})$ *with*

$$\tilde{q}_{i,n} = \begin{cases} \frac{1}{\#A(\tilde{S}_n)} & \text{if} \quad i \in A(\tilde{S}_n) \\ 0 & \text{if} \quad i \notin A(\tilde{S}_n) \end{cases}$$

is asymptotically optimal in the proportional hazards model (5) with conditionally independent censoring.

Proof: It is sufficient (see WIENKE (1996), inequation 4.38) to estimate the expression $\boldsymbol{E}(\Lambda_N(T \mid C) - \Lambda(T \mid C))^2$ to show the convergence to zero for $N \to \infty$. Furthermore,

$$\int\limits_0^t \frac{A(\mathrm{d}x, c)}{\varepsilon_N + B(x, c)} \le \int\limits_0^t \frac{A(\mathrm{d}x, c)}{\varepsilon_{N+1} + B(x, c)} \le \int\limits_0^t \frac{A(\mathrm{d}x, c)}{B(x, c)} \qquad \forall N$$

and with $T = \min\{X, Y\}$

$$\begin{aligned} \boldsymbol{E}\left(\int\limits_0^T \frac{A(dx, C)}{B(x, C)}\right)^2 &= \boldsymbol{E}\Lambda^2(T \mid C) \\ &= \boldsymbol{E} \ln^2 (1 - F_{\vartheta_0}(T \mid C)) \\ &\le \boldsymbol{E} \ln^2 (1 - F_{\vartheta_0}(X \mid C)) \\ &= \int\int \ln^2 (1 - F_{\vartheta_0}(x \mid c))\, P_{\vartheta_0}(\mathrm{d}x \mid c)\, P_C(\mathrm{d}c) = 2\,. \end{aligned}$$

Consequently, the Theorem on Monotone Convergence (SHIRYAYEV, 1984) implies

$$\begin{aligned} &\lim_{N \to \infty} \boldsymbol{E}(\Lambda_N(T \mid C) - \Lambda(T \mid C))^2 \\ &= \lim_{N \to \infty} \boldsymbol{E}\left(\int\limits_0^T \frac{A(\mathrm{d}x, C)}{\varepsilon_N + B(x, C)} - \int\limits_0^T \frac{A(\mathrm{d}x, C)}{B(x, C)}\right)^2 = 0\,. \end{aligned}$$

■

To state the main result some regularity assumptions are necessary.

Assumption (A3): $\int x^2 K(x)\,\mathrm{d}x < \infty$, $\int xK(x)\,\mathrm{d}x = 0$, $\sup_x K(x) = K_{\max} < \infty$ and $\int K(x)\,\mathrm{d}x = 1$.

Assumption (A4): Let $p_C(c)$, $F(t \mid c)$ and $G(t \mid c)$ be twice differentiable w.r.t. c with bounded derivations for all t. Especially assume $p_C(c) < C_{\max} < \infty$.

Lemma 1: *Assume that the conditions* **(A3)** *and* **(A4)** *are fulfilled. Then there exist suitable constants L_1 and L_2 with*

$$\boldsymbol{E}(A_N(T,C) - A(T,C))^2 \le L_1\left(b_N^4 + \frac{1}{Nb_N^2}\right)$$

and

$$\boldsymbol{E}(B_N(T,C) - B(T,C))^2 \le L_2\left(b_N^4 + \frac{1}{Nb_N^2}\right).$$

Proof: Let $(T,C) = (T_{11}, C_{11})$ and L be a suitable constant. Then

$$\begin{aligned}&\boldsymbol{E}(A_N(T_{11},C_{11}) - A(T_{11},C_{11}))^2\\ &\quad = \boldsymbol{E}(\boldsymbol{E}((A_N(T_{11},C_{11}) - A(T_{11},C_{11}))^2 \mid (T_{11},C_{11}))) = \boldsymbol{E}g_N(T_{11},C_{11})\end{aligned}$$

with

$$\begin{aligned}g_N(t,c) &= \boldsymbol{E}\left(\frac{1}{N}\sum_{(i,j)\neq(1,1)} 1(T_{ij} < t, \Delta_{ij} = 1)\, K_b(C_{ij},c) - A(t,c)\right)^2\\
&\le 2\boldsymbol{V}\left(\frac{1}{N}\sum_{(i,j)\neq(1,1)} 1(T_{ij} < t, \Delta_{ij} = 1)\, K_b(C_{ij},c)\right)\\
&\quad + 2\Bigg(-\frac{1}{N}\boldsymbol{E}1(T_{11} < t, \Delta_{11} = 1)\, K_b(C_{11},c)\\
&\quad + \boldsymbol{E}1(T_{11} < t, \Delta_{11} = 1)\, K_b(C_{11},c) - A(t,c)\Bigg)^2\\
&\le \frac{L}{Nb_N^2} + \frac{L}{N^2 b_N^2} + Lb_N^4\\
&\le L_1\left(b_N^4 + \frac{1}{Nb_N^2}\right)\end{aligned}$$

as a consequence of Proposition 3.9 of RÜSCHENDORF (1988) and the first statement follows. The proof of the second one is analogous. ■

Theorem 4: *Let the assumptions* **(A2)**, **(A3)** *and* **(A4)** *be fulfilled,* $b_N = N^{-\frac{1}{6}}$, $\varepsilon_N^{-1} = o(N^{\frac{1}{12}})$ *and* $\hat{S}_n := (\hat{S}_{1,n}(T_1,\Delta_1,C_1),\ldots,\hat{S}_{k,n}(T_k,\Delta_k,C_k))$ *with*

$$\hat{S}_{i,n}(T_i,\Delta_i,C_i) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\Delta_{ij} - \hat{\Lambda}_N(T_{ij} \mid C_{ij})) \tag{15}$$

for $i = 1, \ldots, k$. Then the selection rule $\hat{q}_n = (\hat{q}_{1,n}, \ldots, \hat{q}_{k,n})$

$$\hat{q}_{i,n} = \begin{cases} \dfrac{1}{\#A(\hat{S}_n)} & \text{if} \quad i \in A(\hat{S}_n) \\ 0 & \text{if} \quad i \notin A(\hat{S}_n) \end{cases}$$

with A from (3) is asymptotically optimal for the proportional hazards model (5) with conditionally independent censoring.

5. Simulation Experiments

As Theorem 4 gives only asymptotic results, simulation studies are needed. For this purpose a PASCAL program was written by the author. First we have to specify a concrete model. Therefore we fix the distribution of the covariate C by its distribution function $F_C(x) = 1 - \mathrm{e}^{-x-\frac{1}{2}x^2}$ (Rayleigh-distribution). We calculate lifetime data from three different populations ($k = 3$) with distribution functions $F_{\vartheta_i}(x \mid c) = 1 - \mathrm{e}^{-(1+8(1-\mathrm{e}^{-c}))\,\vartheta_i x}$ (see (5)) and censoring times using $G(x \mid c) = 1 - \mathrm{e}^{-3.33(1+8(1-\mathrm{e}^{-c}))\,x}$. This implies censoring of 25 percent of the lifetimes. Furthermore, let $\vartheta_1 = \vartheta_2 = 10 + \dfrac{10}{\sqrt{n}}$ and $\vartheta_3 = 10$. Consequently, π_3 represents the best population. The kernel in (11) and (12) was chosen by

$$K(x) = \begin{cases} -\dfrac{3}{4}\,x^2 + \dfrac{3}{4} & \text{if} \quad x \in [-1, 1]\,, \\ 0 & \text{if} \quad x \notin [-1, 1]\,. \end{cases}$$

The simulations are based on the following steps:

1. simulation of the covariates C_{ij} (using uniformly distributed random variables on the interval $(0, 1)$ generated by a standard PASCAL routine),
2. simulation of lifetimes X_{ij} and censoring times Y_{ij} conditional on C_{ij},
3. calculation of $(T_{ij}, \Delta_{ij}, C_{ij})$ as a function of (X_{ij}, Y_{ij}, C_{ij}),
4. calculation of statistic (15) and counting events of correct selection.

The following table gives the probability of correct selection (estimated by simulations) for different sample sizes and selection procedures. The procedure Optimal (using the usually unknown distribution/hazard function in (15)) reached the upper Hajek-Le Cam-bound for growing sample size n. Poor results are given by the procedure Mean which selects the population with the largest mean of observed life/censoring times caused by departure from normal distribution and censoring. The simulation results indicate a slight advantage of the proposed procedure in Theorem 4 (Adaptive) w.r.t. the procedure using the KME (KM) because of the violation of the independent assumption in the censoring model.

Table 2

Probability of correct selection using different methods to estimate the unknown distribution/cumulative hazard function in (15). The first column contains the sample size n of each population. Simulation results are presented for the suggested adaptive procedure of Theorem 4 (Adaptive), the procedure using the Kaplan-Meier-estimator (KM), the procedure which selects the population with the largest mean of observed life/censoring times (Mean) and the procedure with known distribution function (Optimal). The last column (Le Cam) shows the Hajek-Le Cam-bound. Simulation of data with 2×10^4 replications each

n	Adaptive	KM	Mean	Optimal	Le Cam
5	0.512	0.512	0.480	0.550	0.594
15	0.530	0.526	0.496	0.567	
25	0.553	0.542	0.494	0.579	
35	0.557	0.540	0.496	0.580	
50	0.565	0.546	0.495	0.581	
75	0.569	0.551	0.499	0.584	
100	0.566	0.545	0.503	0.587	
150	0.567	0.556	0.508	0.590	
200	0.569	0.547	0.501	0.588	
250	0.570	0.550	0.500	0.584	
350	0.565	0.550	0.505	0.585	
500	0.565	0.555	0.500	0.585	
750	0.569	0.555	0.502	0.597	

6. Discussion

The model considered in this article offers an alternative to the usual independent censoring model and allows an application of results from survival analysis to selection procedures. The lifetimes and censoring times are assumed to be conditional independent. Simulations based on this model show that the KME can produce errors in estimation of survival probabilities. In the present setting, the value of the proportional hazards model with conditional independent censoring may lie in producing a class of survival function estimates so that the effect of incorrectly assuming the independent censoring model can be assessed.

Acknowledgements

The author is grateful to Professor F. Liese (Rostock) for many helpful comments and discussions on an earlier version of this manuscript.

Appendix

Proof of Theorem 4: Similar to the proof of Theorem 3 the convergence of expression $\lim\limits_{N\to\infty} \boldsymbol{E}(\hat{\Lambda}_N(T \mid C) - \Lambda_N(T \mid C))^2 = 0$ is to be shown with $\Lambda_N(T \mid C)$ and $\hat{\Lambda}_N(T \mid C)$ from (14) and (15), respectively. With $\bar{\Lambda}_N(t \mid c) := \int\limits_0^t \frac{A(\mathrm{d}x, c)}{\varepsilon_N + B_N(x, c)}$ the inequation

$$\boldsymbol{E}(\hat{\Lambda}_N(T \mid C) - \Lambda_N(T \mid C))^2 \le 2\boldsymbol{E}(\hat{\Lambda}_N(T \mid C) - \bar{\Lambda}_N(T \mid C))^2 + 2\boldsymbol{E}(\bar{\Lambda}_N(T \mid C) - \Lambda_N(T \mid C))^2 \tag{A.1}$$

holds. To estimate the first expression we remark that

$$\begin{aligned}\hat{\Lambda}_N(t \mid c) - \bar{\Lambda}_N(t \mid c) &= \int\limits_0^t \frac{A_N(\mathrm{d}x, c)}{\varepsilon_N + B_N(x, c)} - \int\limits_0^t \frac{A(\mathrm{d}x, c)}{\varepsilon_N + B_N(x, c)} \\ &= \int\limits_0^t B_N^*(x, c)\, A_N^*(dx, c)\end{aligned} \tag{A.2}$$

with $A_N^*(t, c) = A_N(t, c) - A(t, c)$ and $B_N^*(t, c) = (\varepsilon_N + B_N(t, c))^{-1}$. Integrating-by-parts ($A_N^*(t, c)$ and $B_N^*(t, c)$ are left continuous functions w.r.t. t, see SHIRYAYEV, 1984) implies

$$\int\limits_0^t B_N^*(x, c)\, A_N^*(\mathrm{d}x, c) = A_N^*(t, c)\, B_N^*(t, c) - \int\limits_0^t A_N^*(x, c)\, B_N^*(\mathrm{d}x, c)\,. \tag{A.3}$$

Using $B_N^*(t, c) \le \varepsilon_N^{-1}$ and Lemma 4 it holds that

$$\begin{aligned}\boldsymbol{E}(A_N^*(T, C)\, B_N^*(T, C))^2 &\le \frac{1}{\varepsilon_N^2}\, \boldsymbol{E}(A_N^*(T, C))^2 \\ &\le \frac{1}{\varepsilon_N^2}\, L_1 \left(b_N^4 + \frac{1}{N b_N^2}\right).\end{aligned} \tag{A.4}$$

Now we estimate the expression $\int\limits_0^t A_N^*(x, c)\, B_N^*(\mathrm{d}x, c)$. $B_N^*(t, c)$ is piecewise constant w.r.t. t:

$$\begin{aligned}|B_N^*(t+0, c) - B_N^*(t, c)| &= \left|\frac{1}{\varepsilon_N + B_N(t+0, c)} - \frac{1}{\varepsilon_N + B_N(t, c)}\right| \\ &\le \frac{1}{\varepsilon_N^2}\, |B_N(t+0, c) - B_N(t, c)| \\ &\le \frac{1}{\varepsilon_N^2}\, \frac{K_{\max}}{N b_N}\,.\end{aligned}$$

Consequently

$$\left| \int_0^t A_N^*(x,c)\, B_N^*(\mathrm{d}x,c) \right| \le \frac{K_{\max}}{\varepsilon_N^2 N b_N} \sum_{i=1}^{k} \sum_{j=1}^{n} |A_N^*(T_{ij},c)|$$

and

$$\begin{aligned} \boldsymbol{E}\left(\int_0^T A_N^*(x,C)\, B_N^*(\mathrm{d}x,C) \right)^2 &\le \frac{K_{\max}^2}{\varepsilon_N^4 N^2 b_N^2}\, N \sum_{i=1}^{k} \sum_{j=1}^{n} \boldsymbol{E}(A_N^*(T_{ij},C))^2 \\ &\le \frac{K_{\max}^2}{\varepsilon_N^4 b_N^2}\, L_1 \left(b_N^4 + \frac{1}{N b_N^2} \right). \end{aligned} \tag{A.5}$$

Combining (A.2), (A.3), (A.4) and (A.5) with some constant L_3 it holds that

$$\begin{aligned} \boldsymbol{E}(\hat{\Lambda}_N(T \mid C) - \bar{\Lambda}_N(T \mid C))^2 &\le \frac{L_3}{\varepsilon_N^4 b_N^2} \left(b_N^4 + \frac{1}{N b_N^2} \right) \\ &\le L_3 \left(\frac{o(N^{\frac{1}{12}})}{N^{\frac{1}{12}}} \right)^4 \to 0 \end{aligned}$$

as N tends to infinity. Now we deal with the second expression in (A.1). Schwarz's inequality yields

$$\begin{aligned} \boldsymbol{E}(\bar{\Lambda}_N(T \mid C) - \Lambda_N(T \mid C))^2 &\le \boldsymbol{E}\left(\int_0^T \frac{|B_N(x,C) - B(x,C)|\, A(\mathrm{d}x,C)}{(\varepsilon_N + B(x,C))\,(\varepsilon_N + B_N(x,C))} \right)^2 \\ &\le \frac{1}{\varepsilon_N^4}\, \boldsymbol{E}\left(\int_0^T |B_N(x,C) - B(x,C)|\, A(\mathrm{d}x,C) \right)^2 \\ &\le \frac{1}{\varepsilon_N^4}\, \boldsymbol{E} \int_0^T (B_N(x,C) - B(x,C))^2\, A(\mathrm{d}x,C)\; A(T,C)\,. \end{aligned}$$

Introducing $\tilde{B}_N(t,c) := \frac{1}{N} \sum_{(i,j) \neq (1,1)}^{k} \sum^{n} 1(T_{ij} \ge t)\, K_b(C_{ij},c)$ the difference

$$B_N(t,c) - \tilde{B}_N(t,c) \le \frac{K_{\max}}{N b_N} \to 0 \tag{A.6}$$

is asymptotically negligible.

Using $\varphi_N(t,c) := (\tilde{B}_N(t,c) - B(t,c))^2$ it holds (similar to the proof of Lemma 4) with the notation of conditional expectation

$$
\begin{aligned}
&\boldsymbol{E} \int_0^T (\tilde{B}_N(x,C) - B(x,C))^2 \, A(\mathrm{d}x, C) \\
&= \boldsymbol{E} \iiiint \varphi_N(x,c) \, I_{[0,t)}(x) \, A(\mathrm{d}x, c) \, R_{\vartheta_0}(\mathrm{d}t, \mathrm{d}\delta, \mathrm{d}c) \\
&= \iiint \varphi_N(x,c) \, I_{(x,\infty)}(t) \, A(\mathrm{d}x, c) \, (1 - F_{\vartheta_0}(t \mid c)) \, Q(\mathrm{d}t \mid c) \, P_C(\mathrm{d}c) \\
&\quad + \iiint \varphi_N(x,c) \, I_{(x,\infty)}(t) \, A(\mathrm{d}x, c) \, (1 - G(t \mid c)) \, P_{\vartheta_0}(\mathrm{d}t \mid c) \, P_C(\mathrm{d}c) \\
&\leq \boldsymbol{E} \iiint \varphi_N(x,c) \, I_{(x,\infty)}(t) \, p_C(c) \, (1 - G(x \mid c)) \, P_{\vartheta_0}(\mathrm{d}x \mid c) \, Q(\mathrm{d}t \mid c) \, P_C(\mathrm{d}c) \\
&\quad + \boldsymbol{E} \iiint \varphi_N(x,c) \, I_{(x,\infty)}(t) \, p_C(c) \, (1 - G(x \mid c)) \, P_{\vartheta_0}(\mathrm{d}x \mid c) \, P_{\vartheta_0}(\mathrm{d}t \mid c) \, P_C(\mathrm{d}c) \\
&= 2\boldsymbol{E} \iint \varphi_N(x,c) \, p_C(c) \, (1 - G(x \mid c)) \, P_{\vartheta_0}(\mathrm{d}x \mid c) \, P_C(\mathrm{d}c) \\
&\leq 2\boldsymbol{E}(\tilde{B}_N(T,C) - B(T,C))^2 p_C(C) \,.
\end{aligned}
$$

Using relation (A.6) it follow that

$$
\begin{aligned}
&\boldsymbol{E}(\bar{\Lambda}_N(T \mid C) - \Lambda_N(T \mid C))^2 \\
&\leq \frac{2C_{\max}}{\varepsilon_N^4} \, \boldsymbol{E} \int_0^T g(B_N(x,C) - \tilde{B}_N(x,C))^2 \, A(\mathrm{d}x, C) \\
&\quad + \frac{2C_{\max}}{\varepsilon_N^4} \, \boldsymbol{E} \int_0^T (\tilde{B}_N(x,C) - B(x,C))^2 \, A(\mathrm{d}x, C) \\
&\leq \frac{2}{\varepsilon_N^4} \, \frac{K_{\max}^2 C_{\max}^2}{N^2 b_N^2} + \frac{4}{\varepsilon_N^4} \, C_{\max}^2 \boldsymbol{E}(\tilde{B}_N(T,C) - B(T,C))^2 \\
&\leq \frac{L_4}{\varepsilon_N^4 N^2 b_N^2} + \frac{L_4}{\varepsilon_N^4} \, \boldsymbol{E}(B_N(T,C) - B(T,C))^2 \\
&\leq \frac{L_4}{\varepsilon_N^4 N^2 b_N^2} + \frac{L_4}{\varepsilon_N^4} \left(b_N^4 + \frac{1}{N b_N^2} \right) \\
&= \frac{L_4}{N^{\frac{4}{3}}} \left(1 + \frac{o(N^{\frac{1}{12}})}{N^{\frac{1}{12}}} \right)^4 \to 0
\end{aligned}
$$

as N tends to infinity. This completes the proof. ∎

References

ARNOLD, B. C. and KIM, Y. H., 1995: Conditional Proportional Hazards Models. In: N. P. Jewell, A. C. Kimber, M.-L. T. Lee, and G. A. Whitmore (eds.): *Lifetime Data: Models in Reliability and Survival Analysis.* Kluwer Academic Publisher, 21–28.

BECHHOFER, R. E., 1954: A single sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Stat.* **25**, 16–39.

BERAN, R., 1981: *Nonparametric regression with randomly censored survival data.* Technical Report, University of California, Berkeley, CA.

CHENG, P. E., 1989: Nonparametric estimation of survival curve under dependent censorship. *Journal of Statistical Planning and Inference* **23**, 181–191.

COX, D. R., 1972: Regression models and life tables. *Journal of the Royal Statistical Society Ser. B.* **34**, 187–220.

JANSSEN, A., 1989: Local asymptotic normality for randomly censored models with applications to rank tests. *Statistica Neerlandica* **43**, 109–125.

KAPLAN, E. L and MEIER, P., 1958: Nonparametric estimation from incomplete observations. *Journal Amer. Statist. Assoc.* **53**, 457–481.

LAGAKOS, S. W., 1979: General right censoring and its impact on analysis of survival data. *Biometrics* **35**, 139–156.

LE CAM, L., 1986: *Asymptotic Methods in Statistical Decision Theory.* Springer Verlag, Berlin, Heidelberg, New York.

LIESE, F., 1996: Adaptive selection of the best population. *Journal of Statistical Planning and Inference* **54**, 245–269.

MOESCHENBERGER, M. L. and KLEIN, J. P., 1995: Statistical methods for dependent competing risks. *Lifetime Data Analysis* **1**, 195–204.

RÜSCHENDORF, L., 1988: *Asymptotische Statistik.* B.G. Teubner, Stuttgart.

SHIRYAYEV, A. N., 1984: *Probability.* Springer Verlag, New York.

STRASSER, H., 1985: *Mathematical Theory of Statistics.* de Gruyter, Berlin.

TSIATIS, A. A., 1975: A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* (USA) **72**, 20–22.

WIENKE, A., 1996: *Asymptotisch optimale adaptive Auswahlverfahren in semiparametrischen Modellen.* Verlag Hans Jacobs, Lage (Ph.D. thesis).

WITTING, H., 1985: *Mathematische Statistik I.* Teubner, Stuttgart.

Dr. ANDREAS WIENKE
Max Planck Institute for Demographic Research
Doberaner Straße 114
18057 Rostock
Germany

Received, March 1998
Revised, October 1998
Accepted, October 1998