Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Statistical Planning and Inference 136 (2006) 1588–1607

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Likelihood ratio tests for and against ordering of the cumulative incidence functions in multiple competing risks and discrete mark variable models

Hammou El Barmi[a,*], Subhash Kochar[b], John Tsimikas[c]

[a] *Department of Statistics and CIS, Baruch College, City University of New York, New York, 10010, USA*
[b] *Department of Mathematics and Statistics, Portland State University Portland, Oregon 97201, USA*
[c] *Center for Biostatistics, Department of Community Medicine and Health Care, University of Connecticut Health Center, Farmington, CT 06030, USA*

Received 1 April 2004; accepted 15 October 2004
Available online 31 August 2005

Abstract

In this paper we consider the problem of testing the equality of r ($r \geqslant 2$) cumulative incidence functions against an ordered alternative, using the likelihood ratio approach. We assume a discrete time framework and obtain maximum likelihood estimators of the r cumulative incidence functions under the restriction that they are uniformly ordered. The asymptotic null distribution of the derived likelihood ratio test statistic for testing the equality of the cumulative incidence functions against the alternative they are uniformly ordered is of the chi-bar square ($\bar{\chi}^2$) type. In addition to applications within the competing risks setting our methods are also applicable to investigating the association between failure time and a discretized or ordinal mark variable that is observed only at time of failure. We give examples in both the competing risks and mark variable settings and discuss details concerning the implementation of our methods.

MSC: 62G05; 60F17; 62G30

Keywords: Multiple competing risks; Cumulative incidence function; Mark variables; Order restricted inference

* Corresponding author. Tel.: +1 646 312 3384; fax: +1 646 312 3351.
E-mail address: hammou_elbarmi@baruch.cuny.edu (H. El Barmi).

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2005.07.001

1. Introduction

In the standard competing risks model, a unit or subject is exposed to several risks at the same time, but the actual failure is attributed to one cause. In studies with mark variables interest lies in exploring the association between a subject's failure time and the level of a mark variable that is measured only when the subject fails. We only observe (T, δ), where T is the time until failure and δ is the cause of failure or the level of the mark variable at the time of failure. Typically, in both situations, statistical inference is based on the *sub-survival functions*,

$$S_i(t) = P(T \geqslant t, \delta = i), \quad i = 1, \ldots, r$$

or the *cumulative incidence* functions (CIF),

$$F_i(t) = P(T \leqslant t, \delta = i), \quad i = 1, \ldots, r.$$

Note that $\sum_{i=1}^{r} S_i(t) = S_T(t)$ and $\sum_{i=1}^{r} F_i(t) = F_T(t)$ where S_T and F_T are the survival function and the distribution function of T, respectively.

An alternative approach is to compare the *cause* (*mark*) *specific hazard rates*, which for continuous failure times are defined by

$$h_i(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leqslant T < t + \Delta t, \quad \delta = i | T \geqslant t), \quad i = 1, \ldots, r,$$

and for discrete T are given by

$$P(T = t, \delta = i | T \geqslant t).$$

The overall hazard rate for time to failure is given by $h(t) = \sum_{i=1}^{r} h_i(t)$. In the continuous case the sub-survival functions and the cumulative incidence functions can be expressed in terms of the cause specific hazard rates by the relations,

$$S_i(t) = \int_t^{\infty} h_i(u) S_T(u) \, \mathrm{d}u, \quad F_i(t) = \int_0^t h_i(u) S_T(u) \, \mathrm{d}u, \tag{1.1}$$

for $i = 1, 2, \ldots, r$. Similar relations can be established for the discrete case.

In many applications within both the competing risks setting and the studies involving mark variables it is of interest to distinguish between the following alternatives: (i) the cumulative incidence functions are equal, (ii) at least one CIF is greater than the others, (iii) the CIFs are ordered according to a prespecified order. For example, one may wish to investigate whether there is any evidence in the data that the CIFs are ordered according to the level of the mark variable. Possible applications where one may be interested in testing this type of association include (a) studies that investigate the relationship between survival time and a quality of life score, (b) studies that relate survival time to accumulated medical costs, (c) AIDS clinical trial studies investigating the association between failure time and the extent of drug-selected genetic evolution between baseline and failure, an example of which is presented in this paper.

In this paper, we consider the problem of testing the null hypothesis,

$$\mathrm{H}_0 : F_1(t) = F_2(t) = \cdots = F_r(t) \quad \text{for } t \geqslant 0, \tag{1.2}$$

against the alternative $\mathrm{H}_1 - \mathrm{H}_0$, where

$$\mathrm{H}_1 : F_1(t) \leqslant F_2(t) \leqslant \cdots \leqslant F_r(t), \quad \text{for } t \geqslant 0. \tag{1.3}$$

We also consider the hypothesis test:

$$\mathrm{H}_1 \text{ versus } \mathrm{H}_2 - \mathrm{H}_1, \tag{1.4}$$

where H_2 imposes no constraints on the cumulative incidence functions, $F_i, i = 1, 2, \ldots, r$.

We note here that H_0 can be expressed in terms of the sub-survival functions, $\mathrm{H}_0: S_1(t) = S_2(t) = \cdots = S_r(t)$, or in terms of the cause (mark) specific hazard rates, $\mathrm{H}_0: h_1(t) = h_2(t) = \cdots = h_r(t)$. However, H_1 in (1.3) is not equivalent to

$$\mathrm{H}_1' : S_1(t) \geqslant S_2(t) \geqslant \cdots \geqslant S_r(t).$$

It is plausible that the cumulative incidence functions are ordered but the corresponding sub-survival functions cross each other and vice versa.

Note that the hypothesis of ordered cumulative incidence functions, H_1, can be expressed as

$$\mathrm{H}_1 : P(\delta = i | T \leqslant t) \leqslant P(\delta = i+1 | T \leqslant t) \quad i = 1, 2, \ldots, r-1 \ \text{ for } t \geqslant 0.$$

In this form $\mathrm{H}_1 - \mathrm{H}_0$ has the interpretation that given that a unit has failed by time t, the conditional probability of its failing from cause $i+1$ (or having a mark variable level equal to $i+1$) is *uniformly* greater than that from cause i (or having a mark variable equal to i).

Several tests are available in the literature for the special case of testing the equality of two competing risks ($r = 2$). These have been referenced in Aly et al. (1994), El Barmi and Kochar (2003) and in the review paper by Kochar (1995).

We note here that Aly et al. (1994) and Sun and Tiwari (1998) consider the problem of testing the null hypothesis, $\mathrm{H}_0 : F_1(t) = F_2(t)$ against the alternatives

$$\mathrm{H}_1 : F_1(t) \leqslant F_2(t), \quad t \geqslant 0,$$

and

$$\mathrm{H}_1' : S_1(t) \geqslant S_2(t), \quad t \geqslant 0,$$

with strict inequality for some t. Kochar et al. (2002) give a class of tests for testing the equality of two cause specific hazard rates and this class contains the test of Aly et al. (1994) as a special case. Carriere and Kochar (2000) assume continuous failure times and obtain a distribution-free test for the problem of testing H_0 against $\mathrm{H}_1' - \mathrm{H}_0$. Lam (1998) proposed a class of distribution-free tests for testing the equality of k cause specific hazard rates. Kulathinal and Gasbarra (2002) considered the problem of testing the equality of cause specific hazard rates corresponding to m competing risks in k groups. El Barmi and Kochar (2002) consider the same problem with discrete failure times and use the likelihood ratio to test H_0 versus $\mathrm{H}_1' - \mathrm{H}_0$.

In this paper we investigate inference based on the cumulative incidence function assuming *discrete* failure times and mark variables. Discrete failure times arise in competing risk and mark variable studies when the recorded times to failure are grouped in intervals. A discrete mark variable can result by grouping a continuous mark variable in intervals or by observing an ordinal categorical variable at time of failure. We note here that for this framework, and within the competing risks context, Dykstra et al. (1995) obtained the nonparametric maximum likelihood estimates (NPMLEs) of the cause specific hazard rates under the ordered alternative and derived the likelihood ratio test statistic for testing the equality hypothesis of the cause specific hazard rates against the ordered alternative.

Besides many applications in the health sciences, our procedure has potential applications in industrial accelerated life tests. While comparing different brands of a component, the components may be tested in series. The components are functioning in the same environment and their times to failure are generally dependent. The system fails as soon as one of the components fails. Our methods allow testing whether components supplied by different suppliers are of the same quality against the ordered alternative, thus leading to early identification of weak components.

In Section 2 we obtain maximum likelihood estimators of the cumulative incidence functions $F_i, i = 1, 2, \ldots, r$, under H_0 as well as under H_1. In Section 3 we derive the likelihood ratio test for testing H_0 versus $\mathrm{H}_1 - \mathrm{H}_0$, and the likelihood ratio test for testing H_1 versus $\mathrm{H}_2 - \mathrm{H}_1$ and obtain their asymptotic null distribution. In Section 4 we present two examples, one from a competing risks study and one from a clinical trial study investigating the association between survival and a mark variable. The more technical details related to the proofs of the theorems behind our results as well as details on the algorithms needed for the computation of our test statistics and their asymptotic p-values are given in the Appendix. Finally, we note that this work is closely related to that of El Barmi and Dykstra (1995) on testing for and against a set of linear inequality constraints in a multinomial setting.

2. Maximum likelihood estimation

Suppose that we have n individuals exposed to r risks and assume the times and causes of failure represent a random sample from (T, δ). Denote the observations by $(T_1, \delta_1), \ldots, (T_n, \delta_n)$.

In this section we obtain nonparametric maximum likelihood estimates of the cumulative incidence functions, $F_i,\ i = 1, 2, \ldots, r$, under H_0, H_1, and H_2.

For the special case, $r = 2$, Peterson (1977) derived the unrestricted generalized nonparametric MLEs of the two sub-survival functions. The generalized NPMLEs put their weights on the set of observations. Similarly it can be shown that for more than two competing risks ($r > 2$), the unrestricted generalized NPMLE of the ith cumulative incidence function, $F_i(t)$, is

$$\hat{F}_i(t) = \frac{\sum_{j=1}^{n} I(T_j \leqslant t, \delta_j = i)}{n}.$$

In this paper we assume that failures occur on the discrete time points $t_1 < t_2 < \cdots < t_k$ ($t_0 = 0$ and $t_{k+1} = \infty$). For $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, k$, let p_{ij} denote the probability of failure from cause i at time t_j and d_{ij} denote the number of failures from cause i at time t_j. Then

$$F_i(t_j) = pr(T \leqslant t_j, \delta = i) = \sum_{l=1}^{j} p_{il}, \tag{2.1}$$

$i = 1, 2, \ldots, r;\ \ j = 1, 2, \ldots, k$. We write the likelihood function as $L_n = \prod_{i=1}^{r} \prod_{j=1}^{k} p_{ij}^{d_{ij}}$, and the corresponding log-likelihood function as

$$\mathscr{L}_n = n \sum_{i=1}^{r} \sum_{j=1}^{k} \hat{p}_{ij} \ln p_{ij}, \tag{2.2}$$

where

$$\hat{p}_{ij} = \frac{\sum_{l=1}^{n} I[T_l = t_j, \delta_l = i]}{n} = \frac{d_{ij}}{n} \tag{2.3}$$

is the usual unrestricted MLE of p_{ij}.

It is easy to show that under $H_0 : F_1 = F_2 = \cdots = F_r$, the restricted maximum likelihood estimate of p_{ij} is given by

$$\hat{p}_{ij}^{(0)} = \frac{\sum_{l=1}^{n} I[T_l = t_j]}{rn} = \overline{d}_{.j}. \tag{2.4}$$

To facilitate the discussion on finding the maximum likelihood estimates of the p_{ij}s under the hypothesis H_1, we first introduce some notation. Note that the restriction $F_u \leqslant F_{u+1}$ implies k constraints. Hence, for each $u \in \{1, 2, \ldots, r-1\}$ define the k constraint matrices

$$x_{i,j}^{(u,s)} = \begin{cases} 1 & \text{if } i = u \text{ and } j = 1, 2, \ldots, s, \\ -1 & \text{if } i = u+1 \text{ and } j = 1, 2, \ldots, s, \quad s \in \{1, 2, \ldots, k\}, \\ 0 & \text{otherwise.} \end{cases}$$

It is easily seen that $F_u \leqslant F_{u+1}$ is equivalent to

$$\sum_{i=1}^{r} \sum_{j=1}^{k} x_{i,j}^{(u,s)} p_{ij} \leqslant 0, \quad s = 1, 2, \ldots, k.$$

Therefore the maximum likelihood estimates of the p_{ij}s under H_1 are the maximizers of the log-likelihood, $\mathscr{L}_n$, in (2.2), subject to the $k \times (r-1)$ constraints

$$\sum_{i=1}^{r} \sum_{j=1}^{k} x_{i,j}^{(u,s)} p_{ij} \leqslant 0, \quad u = 1, 2, \ldots, r-1, \ \ s = 1, 2, \ldots, k. \tag{2.5}$$

The solution to this optimization problem does not exist in a closed form but can be obtained by using an iterative algorithm based on the Fenchel duality (El Barmi and Dykstra, 1994). This algorithm is presented in Appendix A. We will denote the restricted MLEs under H_1 by $\hat{p}_{ij}^{(1)},\ i = 1, \ldots, r;\ \ j = 1, \ldots, k$.

3. Hypotheses testing

Following the discussion and notation introduced in the previous section, write H_0 and H_1 as

$$\mathrm{H}_0 : \sum_{i=1}^{r} \sum_{j=1}^{k} x_{ij}^{(u,s)} p_{ij} = 0, \quad u = 1, 2, \ldots r-1, \quad s = 1, 2, \ldots, k \tag{3.1}$$

and

$$\mathrm{H}_1 : \sum_{i=1}^{r} \sum_{j=1}^{k} x_{ij}^{(u,s)} p_{ij} \leqslant 0, \quad u = 1, 2, \ldots r-1, \quad s = 1, 2, \ldots, k. \tag{3.2}$$

Let $\pi \subset \{(u, s),\ u = 1, 2, \ldots, r-1,\ s = 1, 2, \ldots, k\}$ be the indices that correspond to an arbitrary subset of the $(r-1) \times k$ equality constraints in (3.1) and let d denote its cardinal; i.e. $d = \mathrm{card}(\pi)$.

First, consider testing H_0 against $\mathrm{H}_{1,\pi} - \mathrm{H}_0$ where

$$\mathrm{H}_{1,\pi} : \sum_{i=1}^{k} \sum_{j=1}^{k} x_{ij}^{(u,s)} p_{ij} = \mathbf{0}, \quad (u, s) \in \pi. \tag{3.3}$$

It is clear from Eq. (2.2) that the log-likelihood ratio test statistic for testing H_0 versus $\mathrm{H}_{1,\pi} - \mathrm{H}_0$ is given by

$$T_{01,\pi} = -2n \sum_{i=1}^{r} \sum_{j=1}^{k} \hat{p}_{ij} [\ln(\hat{p}_{ij}^{(0)}) - \ln(\hat{p}_{ij}(\pi))], \tag{3.4}$$

where $\hat{p}_{ij}^{(0)}$ and $\hat{p}_{ij}(\pi)$, $i = 1, \ldots, r$; $j = 1, \ldots, k$, are the MLEs of p_{ij} under H_0 and under $\mathrm{H}_{1,\pi}$, respectively.

It is a fairly standard exercise to show that the asymptotic distribution of $T_{01,\pi}$ is a chi-square distribution. Nevertheless, we give a detailed proof of this, especially since the arguments contained in our proof are crucial in obtaining the asymptotic distributions of the likelihood ratio test statistics for testing (a) H_0 versus $\mathrm{H}_1 - \mathrm{H}_0$, and (b) H_1 versus $\mathrm{H}_2 - \mathrm{H}_1$.

To derive the asymptotic distributions of the log-likelihood ratio statistic in (3.4), we work with the $(rk-1)$ column vector $\mathbf{p} = (p_{11}, \ldots, p_{rk-1})'$ of cell probabilities. Corresponding to this parameterization, let $\hat{\mathbf{p}}$ denote the unrestricted MLE of $\mathbf{p}$. Also let $\hat{\mathbf{p}}^{(0)}$ and $\hat{\mathbf{p}}(\pi)$ denote the MLEs of $\mathbf{p}$ under H_0 and $\mathrm{H}_{1,\pi}$, respectively. Let $\mathbf{B}$ be the $(rk-1) \times (rk-1)$ matrix:

$$\mathbf{B} = \mathrm{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}'. \tag{3.5}$$

When in the above matrix we let $\mathbf{p} = \mathbf{p}_0$, the true value of $\mathbf{p}$, we obtain $\mathbf{B}^0 = \mathrm{diag}(\mathbf{p}_0) - \mathbf{p}_0\mathbf{p}_0'$, the asymptotic covariance matrix of $\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}_0)$.

Let $\mathbf{H}$ be an $(rk-1) \times (r-1)k$ matrix given by

$$\mathbf{H} = [x_{ij}^{(u,s)} - x_{rk}^{(u,s)}]_{1 \leqslant u \leqslant r-1, 1 \leqslant s \leqslant k, 1 \leqslant i \leqslant r, 1 \leqslant j \leqslant k, (i,j) \neq (r,k)}.$$

We note here that the columns of $\mathbf{H}$ define the $(r-1)k$ order constraints implied by H_1. That is, we can write H_0 and H_1 in Eqs. (3.1) and (3.2) as

$$\mathrm{H}_0 : \mathbf{H}'\mathbf{p} = \mathbf{e} \tag{3.6}$$

and

$$\mathrm{H}_1 : \mathbf{H}'\mathbf{p} \leqslant \mathbf{e}, \tag{3.7}$$

where $\mathbf{e}$ is an $(r-1)k \times 1$ column vector with zeroes everywhere except for the last element which is equal to 1. For an illustration of the matrix $\mathbf{H}$ see Examples 1 and 2.

Let $\mathbf{H}(\pi)$ be the $(rk-1) \times d$ submatrix made of columns of $\mathbf{H}$ whose (u, s) index is in the set π (recall that $d = \mathrm{card}(\pi)$). Clearly the hypothesis $\mathrm{H}_{1,\pi}$ in Eq. (3.3) can be expressed as

$$\mathrm{H}_{1,\pi} : \mathbf{H}'(\pi)\mathbf{p} = \mathbf{e}^*, \tag{3.8}$$

where $\mathbf{e}^*$ is the appropriate subvector of $\mathbf{e}$ in (3.7). The quantity $\hat{\mathbf{p}}(\pi)$, the MLE of $\mathbf{p}$ under $\mathrm{H}_{1,\pi}$, is the maximizer of the log-likelihood in (2.2) subject to the equality constraints in (3.3) or equivalently in (3.8). It can be obtained using the El Barmi and Dykstra (1994) algorithm, shown in Appendix A.

Let the $d \times 1$ column vector $\boldsymbol{\alpha}^*(\pi) = [\alpha_1^*, \ldots, \alpha_d^*]'$ contain the Lagrange multipliers corresponding to the maximization of (2.2) subject to (3.8). Define the $(r-1)k \times (r-1)k$ matrix $\mathbf{R}$, the $(rk-1) \times (rk-1)$ matrices $\mathbf{P}$ and $\mathbf{P}(\pi)$ and the $d \times d$ matrix $\mathbf{R}(\pi)$ by

$$\begin{aligned}
&\mathbf{R} = (\mathbf{H}'\mathbf{B}\mathbf{H})^{-1},\\
&\mathbf{R}(\pi) = (\mathbf{H}'(\pi)\mathbf{B}\mathbf{H}(\pi))^{-1},\\
&\mathbf{P} = \mathbf{B} - \mathbf{B}\mathbf{H}(\mathbf{H}'\mathbf{B}\mathbf{H})^{-1}\mathbf{H}'\mathbf{B},\\
&\mathbf{P}(\pi) = \mathbf{B} - \mathbf{B}\mathbf{H}(\pi)(\mathbf{H}'(\pi)\mathbf{B}\mathbf{H}(\pi))^{-1}\mathbf{H}'(\pi)\mathbf{B}.
\end{aligned} \tag{3.9}$$

Let $\mathbf{B}^0$, $\mathbf{R}^0$, $\mathbf{R}^0(\pi)$, $\mathbf{P}^0$, $\mathbf{P}^0(\pi)$ denote the values of the matrices in (3.5) and (3.9) when evaluated at $\mathbf{p} = \mathbf{p}_0$, where $\mathbf{p}_0$ is the true value of $\mathbf{p}$. It is shown in Appendix B that, under H_0,

$$\sqrt{n}(\hat{\mathbf{p}}^{(0)} - \mathbf{p}_0, \hat{\mathbf{p}}(\pi) - \mathbf{p}_0, \boldsymbol{\alpha}^*(\pi)) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^0(\pi)), \tag{3.10}$$

where the variance-covariance matrix is given by

$$\mathbf{V}^0(\pi) = \begin{bmatrix} \mathbf{P}^0 & \mathbf{P}^0 & \mathbf{0} \\ \mathbf{P}^0 & \mathbf{P}^0(\pi) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}^0(\pi) \end{bmatrix}.$$

The following theorem gives the asymptotic null distribution of $T_{01,\pi}$, the log-likelihood ratio test statistic in (3.4).

Theorem 3.1. 1. *Under* H_0,

$$\sqrt{n}[\hat{\mathbf{p}}^{(0)} - \hat{\mathbf{p}}(\pi)] \xrightarrow{d} N(\mathbf{0}, \mathbf{P}^0 - \mathbf{P}^0(\pi)),$$

2. *Let* $T_{01,\pi}$ *be the log-likelihood ratio test statistic for testing* H_0 *versus* $\mathrm{H}_{1,\pi} - \mathrm{H}_0$. *Let* $d = card(\pi)$ *denote the cardinal of* π. *Under* H_0, *we have*

$$T_{01,\pi} \xrightarrow{d} \chi^2_{k(r-1)-d}.$$

Proof. The proof is given in Appendix B. □

We now consider testing the two hypotheses tests of H_0 versus $\mathrm{H}_1 - \mathrm{H}_0$ and H_1 versus $\mathrm{H}_2 - \mathrm{H}_1$.

Let $\hat{\mathbf{p}}^{(1)}$ denote the restricted MLE of $\mathbf{p}$ under H_1; i.e. under the constraints in (3.7), and let

$$T_{01} = -2n \sum_{i=1}^{r} \sum_{j=1}^{k} \hat{p}_{ij} [\ln(\hat{p}^{(0)}_{ij}) - \ln(\hat{p}^{(1)}_{ij})],$$
$$T_{12} = -2n \sum_{i=1}^{r} \sum_{j=1}^{k} \hat{p}_{ij} [\ln(\hat{p}^{(1)}_{ij}) - \ln(\hat{p}_{ij})]$$

denote the log-likelihood ratio test statistics for testing H_0 versus $\mathrm{H}_1 - \mathrm{H}_0$ and H_1 versus $\mathrm{H}_2 - \mathrm{H}_1$, respectively.

For any positive definite matrix $\mathbf{W}$, define $Q(\mathbf{W})$ as the upper quadrant Gaussian probability,

$$Q(\mathbf{W}) = P(N(\mathbf{0}, \mathbf{W}) > \mathbf{0}), \tag{3.11}$$

and let

$$\begin{aligned}
a_0(\mathbf{p}) &= Q(\mathbf{H}'\mathbf{B}\mathbf{H}) = Q(\mathbf{R}^{-1}),\\
a_d(\mathbf{p}) &= \sum_{\pi, \mathrm{card}(\pi)=d} Q(\mathbf{R}(\pi)) Q(\mathbf{R}^{-1}(\pi^c) - \mathbf{H}'(\pi^c)\mathbf{B}\mathbf{H}(\pi)\mathbf{R}^{-1}(\pi)\mathbf{H}'(\pi)\mathbf{B}\mathbf{H}(\pi^c)),\\
&\quad d = 1, \ldots, k(r-1) - 1,\\
a_{k(r-1)}(\mathbf{p}) &= Q(\mathbf{R}) = 1 - \sum_{d=0}^{k(r-1)-1} a_d,
\end{aligned} \tag{3.12}$$

where π^c denotes the complement of π. $\mathbf{H}(\pi)$ ($\mathbf{H}(\pi^c)$) is the submatrix of $\mathbf{H}$ with the columns determined by the indices in π (π^c).

The following theorem gives the joint asymptotic distribution of (T_{01}, T_{12}), under H_0.

Theorem 3.2. *Under* H_0 *and for any* $t_1 > 0$ *and* $t_2 > 0$, *we have*

$$\lim_{n\to\infty} P(T_{01} \geqslant t_1, T_{12} \geqslant t_2) = \sum_{d=0}^{k(r-1)} a_d(\mathbf{p}_0) P(\chi^2_{k(r-1)-d} \geqslant t_1) P(\chi^2_d \geqslant t_2) \tag{3.13}$$

with $\chi^2_0 \equiv 0$.

Proof. The proof is given in Appendix C. □

In particular, the null asymptotic distribution of the log-likelihood ratio test statistic for testing H_0 versus $\mathrm{H}_1 - \mathrm{H}_0$ is obtained by

$$\lim_{n\to\infty} P(T_{01} \geqslant t) = \sum_{d=0}^{k(r-1)} a_d(\mathbf{p}_0) P(\chi^2_{k(r-1)-d} \geqslant t) \tag{3.14}$$

with $\chi^2_0 \equiv 0$. Similarly the asymptotic distribution of the log-likelihood ratio test statistic for testing H_1 versus $\mathrm{H}_2 - \mathrm{H}_1$, under the hypothesis of equality of the cumulative incidence functions, is obtained by

$$\lim_{n\to\infty} P(T_{12} \geqslant t) = \sum_{d=0}^{k(r-1)} a_d(\mathbf{p}_0) P(\chi^2_d \geqslant t) \tag{3.15}$$

with $\chi^2_0 \equiv 0$. In practice, since $\mathbf{p}_0$, the true value of $\mathbf{p}$, is unknown, the weights $a_d(\mathbf{p}_0)$, as defined in (3.12), are estimated by $a_d(\hat{\mathbf{p}}^{(0)})$, their consistent estimators under H_0. That is, let $\hat{\mathbf{B}}^{(0)}$ be the estimated covariance matrix of $\hat{\mathbf{p}}$, under H_0, obtained by setting $\mathbf{p} = \hat{\mathbf{p}}^{(0)}$ in Eq. (3.5). As indicated by (3.12), computation of the estimated asymptotic p-values rests in obtaining the weights, $a_d(\hat{\mathbf{p}}^{(0)})$, each of which involves estimation of multiple multivariate quadrant probabilities, defined in (3.11). These can be efficiently obtained after successive applications of the Sweep operator to the matrix $\mathbf{H}'\hat{\mathbf{B}}^{(0)}\mathbf{H}$ combined with a routine for approximating Gaussian quadrant probabilities. The matrix $\mathbf{H}'\hat{\mathbf{B}}^{(0)}\mathbf{H}$ involves the sample cumulative frequency of failures and is given in the examples. Details on the efficient estimation of the weights are given in Appendix D.

If $r=2$ which is the case discussed in El Barmi and Kochar (2002), $a_d(\mathbf{p})=p(d,k,\mathbf{p}_r)$, $d=0,1,\ldots,k$, where $p(0,k,\mathbf{p}_r)$ is the probability that $E_{\mathbf{p}_r}[\mathbf{U}|\mathscr{I}]$ is identically zero and $p(d,k,\mathbf{p}_r)$, $d=1,2,\ldots,k$, is the probability that $E_{\mathbf{p}_r}[\mathbf{U}|\mathscr{I}]$ has d distinct values. Here $\mathbf{p}_r=(p_{21},\ldots,p_{2k})'$, $\mathbf{U}=(U_1,U_2,\ldots,U_k)'$ where U_is are independent and U_i has a normal distribution with mean 0 and variance $1/p_{2i}$ and $E_{\mathbf{p}_r}[\mathbf{U}|\mathscr{I}]$ is the least squares projection of $\mathbf{U}$ onto $\mathscr{I}=\{\mathbf{x}\in\mathscr{R}^k, 0\geqslant x_1\geqslant x_2\geqslant\cdots\geqslant x_k\}$. So that for testing H_0 against $\mathrm{H}_1-\mathrm{H}_0$, if there is evidence that $p_{21}, p_{22},\ldots,p_{2k}$ do not vary too much, a test based on equal weights critical value will have a significance level reasonably close to the reported value. These equal weights level probabilities can be found in Robertson et al. (1988). Since we have 0 as an upper bound in the cone $\mathscr{I}$, the value k should be increased by 1 to account for it. As pointed out in El Barmi and Kochar (2002), this is like having $k+1$ normal means indexed by $0,1,2,\ldots,k$ with the weight associated with the variable indexed by 0 being ∞. Finally, they also showed that

$$\sup_{\mathbf{p}\in\mathrm{H}_0}\lim_{n\to\infty} P(T_{01} \geqslant t) = \tfrac{1}{2}[P(\chi^2_{k-1}\geqslant t) + P(\chi^2_k \geqslant t)] \tag{3.16}$$

and

$$\begin{aligned}\sup_{\mathbf{p}\in\mathrm{H}_1}\lim_{n\to\infty} P(T_{12} \geqslant t) &= \sup_{\mathbf{p}\in\mathrm{H}_0}\lim_{n\to\infty} P(T_{12} \geqslant t)\\ &= \sum_{d=1}^{k+1}\binom{k}{d-1}2^{-k}P(\chi^2_{d-1}\geqslant t).\end{aligned} \tag{3.17}$$

We have not being able to extend these results to $r > 2$ but from the well known properties of the weights of a chi-bar square distribution (3.16) always hold with $\leqslant$ instead of $=$.

4. Examples

Example 1. For our first illustration we consider the mortality data on RFM strain male mice reported in Hoel (1972). Two risks are considered. The second risk is cancer and the first combines all other risks. The failure times are grouped into $k = 6$ categories. Thus, we have two competing risks, $r = 2$, and $k = 6$ time periods. In this case the constraints are

$$\sum_{j=1}^{s} p_{1j} \leqslant \sum_{j=1}^{s} p_{2j}, \quad s = 1, \ldots, 6,$$

and the matrix $\mathbf{H}$ is an 11×6 matrix and is given by

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ -1 & -1 & -1 & -1 & -1 & 0 \\ 0 & -1 & -1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}.$$

Under $\mathrm{H}_0 : \sum_{j=1}^{s} p_{1j} = \sum_{j=1}^{s} p_{2j}, \ s = 1, 2, \ldots, 6$, we can write the vector $\mathbf{p}_0$ as

$$\mathbf{p}_0 = [p_1, p_2, p_3, p_4, p_5, p_6, p_1, p_2, p_3, p_4, p_5]'.$$

It is easy to show that

$$\mathbf{H}'\mathbf{B}^0\mathbf{H} = 2\begin{bmatrix} p_1 & p_1 & p_1 & p_1 & p_1 & p_1 \\ p_1 & (p_1+p_2) & (p_1+p_2) & (p_1+p_2) & (p_1+p_2) & (p_{11}+p_{12}) \\ p_1 & (p_1+p_2) & (p_1+p_2+p_3) & (p_1+p_2+p_3) & (p_1+p_2+p_3) & (p_1+p_2+p_3) \\ p_1 & (p_1+p_2) & (p_1+p_2+p_3) & (p_1+p_2+p_3+p_4) & (p_1+p_2+p_3+p_4) & (p_1+p_2+p_3+p_4) \\ p_1 & (p_1+p_2) & (p_1+p_2+p_3) & (p_1+p_2+p_3+p_4) & (p_1+p_2+p_3+p_4+p_5) & (p_1+p_2+p_3+p_4+p_5) \\ p_1 & (p_1+p_2) & (p_1+p_2+p_3) & (p_1+p_2+p_3+p_4) & (p_1+p_2+p_3+p_4+p_5) & \frac{1}{2} \end{bmatrix}.$$

The data and estimates are shown in Table 1. In this table the column labeled $d_{.j}$ is the total number of failures at time j, combined over both causes, and the column labeled $D_{.j}$ contains the corresponding cumulative counts.

Table 1

No.	Interval	d_{1j}	d_{2j}	$d_{.j}$	$D_{.j}$	$\hat{p}_{1j}$	$\hat{p}_{2j}$	$\hat{p}_{1j}^{(1)}$	$\hat{p}_{2j}^{(1)}$
1	(0, 350)	15	18	33	33	0.1515	0.1818	0.1515	0.1818
2	[350, 450)	6	7	13	46	0.0606	0.0707	0.0606	0.0707
3	[450, 550)	6	4	10	56	0.0606	0.0404	0.0606	0.0404
4	[550, 650)	8	18	26	82	0.0808	0.1818	0.0808	0.1818
5	[650, 750)	2	12	14	96	0.0202	0.1212	0.0202	0.1212
6	[750, 850)	2	1	3	99	0.0202	0.0101	0.0152	0.0152

The matrix, $\mathbf{H}'\hat{\mathbf{B}}^{(0)}\mathbf{H}$, needed for estimation of the weights in (3.13)–(3.15), is given by

$$\mathbf{H}'\hat{\mathbf{B}}^{(0)}\mathbf{H} = \frac{1}{99}\begin{bmatrix} 33 & 33 & 33 & 33 & 33 & 33 \\ 33 & 46 & 46 & 46 & 46 & 46 \\ 33 & 46 & 56 & 56 & 56 & 56 \\ 33 & 46 & 56 & 82 & 82 & 82 \\ 33 & 46 & 56 & 82 & 96 & 96 \\ 33 & 46 & 56 & 82 & 96 & 99 \end{bmatrix}.$$

The estimated weights needed for the null asymptotic distribution of the test statistic T_{01}, are given below

$$a_0(\hat{\mathbf{p}}^{(0)}) = 0.2775879, \quad a_1(\hat{\mathbf{p}}^{(0)}) = 0.4532581, \quad a_2(\hat{\mathbf{p}}^{(0)}) = 0.2177982,$$
$$a_3(\hat{\mathbf{p}}^{(0)}) = 0.0403585, \quad a_4(\hat{\mathbf{p}}^{(0)}) = 0.0107544, \quad a_5(\hat{\mathbf{p}}^{(0)}) = 0.0002417,$$
$$a_6(\hat{\mathbf{p}}^{(0)}) = 0.0000012.$$

For this example the value of $T_{01} = 12.6247$ and the value of $T_{12} = 0.3397$.

The estimated approximate p-value for testing H_0 vs $\mathrm{H}_1 - \mathrm{H}_0$ is $p_{\mathrm{val}} = 0.02915$. The estimated approximate p-value for testing H_1 vs $\mathrm{H}_2 - \mathrm{H}_1$ is $p_{\mathrm{val}} = 0.7961356$.

Example 2. In our second illustration we consider data from a randomized study conducted by the Adult AIDS Clinical Trials Group (AACTG) to evaluate two combination antiretroviral treatments in terms of their ability to suppress HIV viral load. The failure time, T, was defined as the time from randomization until plasma HIV levels rose above 1000 copies/ml. At failure a measure of acquired mutational distance during the trial was obtained. This distance is a measure of the accumulated HIV genetic resistance due to treatment exposure and is only obtained when a subject fails. Gilbert et al. (2004) normalize this distance so that it lies in the interval [0, 1]. For our purposes we discretize the normalized distance measure, call it V, and consider $r=3$ groups. A subject is classified as belonging to group 1 if $V \in (0, 1/3]$, to group 2 if $V \in (1/3, 2/3]$ and to group 3 if $V \in (2/3, 1]$. Also we consider $k=3$ failure time intervals. We take $j=1$ if $T \in (0, 5]$, $j=2$ if $T \in (5, 20]$ and $j=3$ if $T \in (20, 50]$. The data is given in Table 2. Hence we have $r=3$, and $k=3$. The matrix $\mathbf{H}$ is an 8×6 matrix

Table 2

Interval	d_{1j}	d_{2j}	d_{3j}	$d_{.j}$	$D_{.j}$	$\hat{p}_{1j}$	$\hat{p}_{2j}$	$\hat{p}_{3j}$	$\hat{p}_{1j}^{(1)}$	$\hat{p}_{2j}^{(1)}$	$\hat{p}_{3j}^{(1)}$
(0, 5]	5	7	7	19	19	0.1111	0.1556	0.1556	0.1111	0.1458	0.1667
(5, 20]	6	5	4	15	34	0.1333	0.1111	0.0889	0.1333	0.1042	0.0952
(20, 50]	4	4	3	11	45	0.0889	0.0889	0.0667	0.0889	0.0833	0.0714

and is given by

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ -1 & -1 & -1 & 1 & 1 & 2 \\ 0 & -1 & -1 & 0 & 1 & 2 \\ 0 & 0 & -1 & 0 & 0 & 2 \\ 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}.$$

In the setting of this example, H_0 states no association between the cumulative risk function and the level of V, the acquired mutational distance. The hypothesis H_1 states that for every time period the cumulative incidence increases as the level of V increases. Under H_0, we can write the vector $\mathbf{p}_0$ as

$$\mathbf{p}_0 = [p_1, p_2, p_3, p_1, p_2, p_3, p_1, p_2]'.$$

The matrix, $\mathbf{H}'\hat{\mathbf{B}}^{(0)}\mathbf{H}$, needed for estimation of the weights in (3.14), is given by

$$\mathbf{H}'\hat{\mathbf{B}}^{(0)}\mathbf{H} = \frac{3}{45}\begin{bmatrix} 19 & 19 & 19 & -9.5 & -9.5 & -9.5 \\ 19 & 34 & 34 & -9.5 & -17 & -17 \\ 19 & 34 & 45 & -9.5 & -17 & -22.5 \\ -9.5 & -9.5 & -9.5 & 19 & 19 & 19 \\ -9.5 & -17 & -17 & 19 & 34 & 34 \\ -9.5 & -17 & -22.5 & 19 & 34 & 45 \end{bmatrix}.$$

The estimated weights needed for the null asymptotic distribution of the test statistic T_{01}, are given below

$$a_0(\hat{\mathbf{p}}^{(0)}) = 0.0516802, \quad a_1(\hat{\mathbf{p}}^{(0)}) = 0.2325504, \quad a_2(\hat{\mathbf{p}}^{(0)}) = 0.3604903,$$
$$a_3(\hat{\mathbf{p}}^{(0)}) = 0.2532694, \quad a_4(\hat{\mathbf{p}}^{(0)}) = 0.0872682, \quad a_5(\hat{\mathbf{p}}^{(0)}) = 0.0143132,$$
$$a_6(\hat{\mathbf{p}}^{(0)}) = 0.0008899.$$

For this example the value of $T_{01} = 0.8958848$ and the value of $T_{12} = 0.1334323$. The estimated approximate p-value for testing H_0 vs $\mathrm{H}_1 - \mathrm{H}_0$ is $p_{\mathrm{val}} = 0.8803156$. The estimated approximate p-value for testing H_1 vs $\mathrm{H}_2 - \mathrm{H}_1$ is $p_{\mathrm{val}} = 0.9647699$. Thus we do not have enough evidence to conclude association between the failure time and the level of the mark variable, a result consistent with the conclusion in Gilbert et al. (2004). Evidently, our test does dependent on how the data are grouped.

Acknowledgements

The authors would like to thank the editor and a referee, for helpful comments and suggestions that led to a much improved paper. Hammou El Barmi thanks the City University of New York for its support through PSC-CUNY Research Award Program.

Appendix A. Description of algorithm

El Barmi and Dykstra (1994) showed that, if $\mathbf{y}^*$ solves

$$\max_{\mathbf{y}\in K_C^*} \sum_{i=1}^{m} \hat{p}_i \ln(1+y_i), \tag{A.1}$$

where $K_C^* = \{\mathbf{y}, \sum_{i=1}^m x_i y_i \leqslant 0, \ \forall \mathbf{x} \in C\}$, for C a closed, convex subset of $\mathscr{P} = \{(x_1, x_2, \ldots, x_m)', x_i \geqslant 0, \forall i, \sum_{i=1}^m x_i = 1\}$, then

$$p_i^* = \frac{\hat{p}_i}{1+y_i^*}, \quad i = 1, 2, \ldots, m,$$

solves

$$\max_{C} \prod_{i=1}^{m} p_i^{\hat{p}_i}. \tag{A.2}$$

Here $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_m)' \in \mathscr{P}$ and is in general the vector of the relative frequencies. In the event that $C = \{\mathbf{p} \in \mathscr{P}, \sum_{i=1}^m p_i a_{ij} = 0, \ j = 1, 2, \ldots, s\}$, a set of linear constraints, it is easy to show that (A.1) reduces to

$$\max_{\alpha_j;\, j=1,\ldots,s} \sum_{i=1}^{m} \hat{p}_i \ln\left(1 + \sum_{j=1}^{s} \alpha_j a_{ij}\right). \tag{A.3}$$

Note that our maximization problem defined by maximizing the log-likelihood in (2.2) subject to the constraints in Eq. (2.5) is of the type (A.2). Let $(\alpha_1^*, \alpha_2^*, \ldots, \alpha_s^*)$ denote the maximizing values of the above expression. Then the solution to the maximization in (A.2) is given by

$$p_i^* = \frac{\hat{p}_i}{1+\sum_{j=1}^{s} \alpha_j^* a_{ij}}, \quad i = 1, \ldots, m. \tag{A.4}$$

The following algorithm can be used to find $(\alpha_1^*, \alpha_2^*, \ldots, \alpha_s^*)$ and hence $p_1^*, p_2^*, \ldots, p_s^*$.

Algorithm.

- *Step* 1: initially $\alpha_j = 0$, $j = 1, 2, \ldots, s$, $v = 1$
- *Step* 2: Find the optimal value of α_v over $\mathscr{R}$ with all the other αs held fix. This value of α_v replaces it previous value.

- If $v < s$ set $v = v + 1$, if $v = s$, set $v = 1$.
- Go to step 2. Find the optimal value of α_v over $\mathscr{R}$ with all the other αs held fix. This value of α_v replace its previous value.
  These steps are repeated for $v = 1, 2, \ldots$ until sufficient accuracy is attained. We note here that $(\alpha_1^*, \alpha_2^*, \ldots, \alpha_s^*)$ are the Lagrange multipliers corresponding to maximizing the log-likelihood function subject to the constraints $\mathbf{p} \in C$, i.e.

$$\sum_{i=1}^{k} a_{ij} p_i = 0, \quad j = 1, 2, \ldots, s.$$

  If it is the case that $C = \{\mathbf{p} \in \mathscr{P}, \sum_{i=1}^{k} p_i a_{ij} \leqslant 0,\ j = 1, 2, \ldots, s\}$, then Step 2 of the algorithm should be replaced by
- *Step* 2*: Find the optimal value of α_v over $\mathscr{R}^+$ with all the other αs held fix. This value of α_v replaces its previous value.

We note that at a given step of the algorithm, the desired α can be found very quickly by a Newton–Raphson (in general 2–3 steps to find the optimum value in each step). This procedure been successfully used by Dykstra et al. (1996) for 60 linear constraints in a 61-dimensional space.

Appendix B. Proof of Theorem 3.1

Using a Taylor expansion, under H_0, we have

$$\begin{aligned} T_{01,\pi} &= -2n \sum_{i=1}^{r} \sum_{j=1}^{k} \hat{p}_{ij} [\ln(\hat{p}_{ij}^{(0)}) - \ln(\hat{p}_{ij}(\pi))] \\ &= n \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{1}{\hat{p}_{ij}(\pi)} (\hat{p}_{ij}(\pi) - \hat{p}_{ij}^{(0)})^2 + \mathrm{o_p}(1). \end{aligned} \tag{B.1}$$

Write the likelihood function as

$$L_n = \left[\prod_{(i,j)\neq(r,k)} p_{ij}^{n_{ij}} \right] \left[1 - \sum_{(i,j)\neq(r,k)} p_{ij} \right]^{n_{rk}}.$$

Let

$$\mathscr{D}\mathscr{L}_n(\mathbf{p}) = \left(\frac{\partial}{\partial p_{ij}} \ln L_n(\mathbf{p}) \right)_{(i,j)\neq(r,k)}$$

be the gradient of the log-likelihood and $\mathbf{p}_0 = (p_{11}^0, p_{12}^0, \ldots, p_{r1}^0, \ldots, p_{r,k-1}^0)' \in \mathrm{H}_0$ be the true value of $\mathbf{p}$. Then we have

$$\begin{aligned}
\sqrt{n}[\hat{\mathbf{p}}^{(0)} - \mathbf{p}_0] &= \frac{1}{\sqrt{n}}\mathbf{P}^0\mathscr{D}\mathscr{L}_n(\mathbf{p}_0) + \mathrm{o_p}(1),\\
\sqrt{n}[\hat{\mathbf{p}}^{(0)}(\pi) - \mathbf{p}^0] &= \frac{1}{\sqrt{n}}\mathbf{P}^0(\pi)\mathscr{D}\mathscr{L}_n(\mathbf{p}_0) + \mathrm{o_p}(1),\\
\sqrt{n}\alpha^*(\pi) &= \frac{1}{\sqrt{n}}\mathbf{Q}^0(\pi)\mathscr{D}\mathscr{L}_n(\mathbf{p}_0) + \mathrm{o_p}(1),
\end{aligned} \tag{B.2}$$

where $\hat{\mathbf{p}}^{(0)} = (\hat{p}_{11}^{(0)}, \hat{p}_{12}^{(0)}, \ldots, \hat{p}^{(0)})_{r,k-1})'$ and $\hat{\mathbf{p}}(\pi) = (\hat{p}_{11}(\pi), \hat{p}_{12}(\pi), \ldots, \hat{p}_{r,k-1}(\pi))'$ are the maximum likelihood estimators of $\mathbf{p}$ under H_0 and $\mathrm{H}_{1,\pi}$, respectively and $\alpha^*(\pi)$ is the Lagrange multiplier associated with the maximization of the likelihood function under $\mathrm{H}_{1,\pi}$. If the $\hat{p}_{ij} > 0$ for all (i, j) then $\hat{\mathbf{p}}^0$ and $\hat{\mathbf{p}}(\pi)$ will be unique. The matrices $\mathbf{P}$ and $\mathbf{P}(\pi)$ are as defined in (3.9) and $\mathbf{Q}(\pi) = -\mathbf{BH}(\mathbf{H}'\mathbf{BH})^{-1}$. $\mathbf{P}^0$, $\mathbf{P}^0(\pi)$ and $\mathbf{Q}^0(\pi)$ are the values of the matrices when $\mathbf{B} = \mathbf{B}(\mathbf{p}_0)$, as defined before with $\mathbf{p} = \mathbf{p}_0$.

Therefore under H_0, we have

$$\begin{aligned}
&\sqrt{n}[\hat{p}_{11}^0 - \hat{p}_{11}(\pi), \hat{p}_{12}^0 - \hat{p}_{12}(\pi), \ldots, \ldots, \hat{p}_{r,k-1}^0 - \hat{p}_{r,k-1}(\pi))'\\
&\quad = \frac{1}{\sqrt{n}}[\mathbf{P}^0 - \mathbf{P}^0(\pi)]\mathscr{D}\ln\mathscr{L}_n(\mathbf{p}_0) + \mathrm{o_p}(1)
\end{aligned}$$

and therefore converges in distribution as n goes to infinity to a multivariate normal distribution with mean vector zero and covariance matrix given by

$$(\mathbf{P}^0 - \mathbf{P}^0(\pi))\mathbf{B}^{-1}(\mathbf{P}^0 - \mathbf{P}^0(\pi)) = \mathbf{P}^0 - \mathbf{P}^0(\pi).$$

Assume without loss of generality assume that $\mathbf{H}(\pi)$ is made of the first d columns of $\mathbf{H}$, then

$$\begin{aligned}
\mathbf{P}^0 - \mathbf{P}^0(\pi) &= \mathbf{BH}(\mathbf{H}'\mathbf{BH})^{-1}\mathbf{H}'\mathbf{B} - \mathbf{BH}(\pi)(\mathbf{H}'(\pi)\mathbf{BH}(\pi))^{-1}\mathbf{H}'(\pi)\mathbf{B}\\
&= (\mathbf{P}^0 - \mathbf{P}^0(\pi))\mathbf{B}^{-1}(\mathbf{P}^0 - \mathbf{P}^0(\pi))\\
&= (\mathbf{BH}(\pi^c) - \mathbf{BH}(\pi)\Sigma_{12})\Sigma^{-1}(\mathbf{H}'(\pi)\mathbf{B} - \Sigma_{12}\mathbf{H}(\pi)'\mathbf{B}),
\end{aligned} \tag{B.3}$$

where

$$\begin{aligned}
\Sigma &= \mathbf{H}'(\pi^c)[\mathbf{B} - \mathbf{BH}(\pi)(\mathbf{H}'(\pi)\mathbf{BH}(\pi))^{-1}\mathbf{H}(\pi)\mathbf{B}]\mathbf{H}(\pi^c),\\
\Sigma_{12} &= \Sigma_{21}' = (\mathbf{H}'(\pi)\mathbf{BH}(\pi))^{-1}\mathbf{H}(\pi)'\mathbf{BH}(\pi^c)
\end{aligned}$$

and $\mathbf{H}(\pi^c)$ is made of the remaining columns of $\mathbf{H}$.

Since $\sqrt{n}(\hat{p}_{11}^0-\hat{p}_{11}(\pi),\ldots,\hat{p}_{r,k-1}^0-\hat{p}_{r,k-1}(\pi))'$ converges in distribution to $N(0,\mathbf{P}^0(\pi)-\mathbf{P}^0)$, it follows from (3.9) that the asymptotic covariance of $\sqrt{n}[\hat{p}_i^0-\hat{p}_i(\pi)]$ and $\sqrt{n}[\hat{p}_j^0-\hat{p}_j(\pi)]$ is

$$\mathbf{v}_i'(\pi)\Sigma^{-1}\mathbf{v}_j(\pi)-\mathbf{v}_i'(\pi)\Sigma^{-1}\Sigma_{21}\mathbf{u}_j(\pi)-\mathbf{u}_i'(\pi)\Sigma_{12}\Sigma^{-1}\mathbf{v}_j(\pi) \\ +\mathbf{u}_i'(\pi)\Sigma_{12}\Sigma^{-1}\Sigma_{21}\mathbf{u}_j(\pi),$$

where $\mathbf{BH}(\pi)=(\mathbf{u}_1(\pi),\mathbf{u}_2(\pi),\ldots,\mathbf{u}_{rk-1}(\pi))'$, $\mathbf{BH}(\pi^c)=(\mathbf{v}_1(\pi),\mathbf{v}_2(\pi),\ldots,\mathbf{v}_{rk-1}(\pi))'$, $\mathbf{u}_{rk}(\pi)=-\sum_{j=1}^{rk-1}\mathbf{u}_j(\pi)$ and $\mathbf{v}_{rk}(\pi)=-\sum_{j=1}^{rk-1}\mathbf{v}_j(\pi)$. Consequently, under H_0

$$\sqrt{n}\left(\frac{\hat{p}_{11}^0-\hat{p}_{11}(\pi)}{\sqrt{p_{11}^0}},\ldots,\frac{\hat{p}_{rk}^0-\hat{p}_{rk}(\pi)}{\sqrt{p_{rk}^0}}\right) \\ \xrightarrow{d} N(\mathbf{0},[M_2'-M_1'\Sigma_{12}]'\Sigma^{-1}[M_2-\Sigma_{21}M_1]),$$

where

$$M_1=\left(\mathbf{u}_1(\pi)\Big/\sqrt{p_{11}^0},\mathbf{u}_2(\pi)\Big/\sqrt{p_{12}^0},\ldots,\mathbf{u}_{rk}(\pi)\Big/\sqrt{p_{kk}^0}\right)',$$
$$M_2=\left(\mathbf{v}_1(\pi)\Big/\sqrt{p_{11}^0},\mathbf{v}_2(\pi)\Big/\sqrt{p_{12}^0},\ldots,\mathbf{v}_{rk}(\pi)\Big/\sqrt{p_{rk}^0}\right)'.$$

It then follows that

$$\sqrt{n}\left(\frac{\hat{p}_{11}^0-\hat{p}_{11}(\pi)}{\sqrt{p_{11}^0}},\ldots,\frac{\hat{p}_{rk}^0-\hat{p}_{rk}(\pi)}{\sqrt{p_{rk}^0}}\right)$$

converges in distribution to

$$(Y_{11},Y_{12},\ldots,Y_{r1},Y_{rk})'=[M_2'-M_1'\Sigma_{12}]'\Sigma^{-1/2}(Z_1,Z_2,\ldots,Z_{(r-1)k-d})',$$

where Z_i are i.i.d $N(0,1)$. It follows that

$$\sum_{i=1}^{r}\sum_{j=1}^{k}n\,\frac{(\hat{p}_{ij}^0-\hat{p}_{ij}(\pi))^2}{p_{ij}^0}$$

converges in distribution to

$$\sum_{i=1}^{r}\sum_{j=1}^{k} Y_{ij}^2 = \mathbf{Z}'\Sigma^{-1/2}[M_2 - \Sigma_{21}M_1][M_2' - M_1'\Sigma_{12}]'\Sigma^{-1/2}\mathbf{Z}$$
$$= \sum_{l=1}^{(r-1)k-d} Z_i^2$$

which has a chi-square distribution with $(r-1)k-d$ degrees of freedom. Here $\mathbf{Z}' = (Z_1, Z_2, \ldots, Z_{(r-1)k-d})$ and the second equality holds because

$$[M_2 - \Sigma_{21}M_1][M_2' - M_1'\Sigma_{12}]' = \Sigma.$$

Appendix C. Proof of Theorem 3.2

Let $\mathscr{C}$ be the set of all subsets of constraints. If follows from El Barmi and Dykstra (1994) (see Appendix A) that $\hat{\mathbf{p}}^{(1)}(\pi)$, the maximizer of

$$\prod_{i=1}^{r}\prod_{j=1}^{k} p_{ij}^{n_{ij}} \tag{C.1}$$

subject to

$$\sum_{i=1}^{r}\sum_{j=1}^{k} p_{ij}x_{ij}^{(u,s)} = 0,\ (u,s) \in \pi,$$

and the solution, $\boldsymbol{\alpha}^*(\pi)$, of the maximization

$$\max_{\alpha_{us},(u,s)\in\pi} \sum_{i=1}^{r}\sum_{j=1}^{k} \hat{p}_{ij} \ln\left(1 + \sum_{(u,s)\in\pi} \alpha_{us}x_{ij}^{(u,s)}\right), \tag{C.2}$$

satisfy

$$\hat{p}_{ij}^{(1)} = \frac{\hat{p}_{ij}}{1 + \sum_{(u,s)\in\pi} \alpha_{us}^*(\pi)x_{ij}^{(u,s)}}, \quad \forall (i,j).$$

Moreover $\hat{p}^{(1)} = \hat{p}^{(1)}(\pi)$ for precisely one π. Also, we have

$$\hat{p}^{(1)} = \hat{p}^{(1)}(\pi) \iff \begin{cases} \alpha_{us}^*(\pi) > 0,\ (u,s) \in \pi, \\ \sum_{i=1}^{r}\sum_{j=1}^{k} x_{ij}^{(u,s)}\hat{p}_{ij}^{(1)}(\pi) \leqslant 0,\ (u,s) \in \pi^c. \end{cases}$$

Combining the above statements gives

$$
\begin{aligned}
&P(T_{01} \geqslant t_1, T_{12} \geqslant t_2) \\
&\quad = \sum_{\pi \in \mathscr{C}} P(T_{01} \geqslant t_1, T_{12} \geqslant t_2, \hat{\mathbf{p}}^{(1)} = \hat{\mathbf{p}}^{(1)}(\pi)) \\
&\quad = \sum_{\pi \in \mathscr{C}} P\left(T_{01} \geqslant t_1, T_{12} \geqslant t_2, \alpha^*_{us}(\pi) > 0, (u,s) \in \pi, \right. \\
&\qquad \left. \sum_{i,j} x_{ij}^{(u,s)} \hat{p}_{ij}^{(1)}(\pi) \leqslant 0, (u,s) \in \pi^c \right) \\
&\quad = \sum_{\pi \in \mathscr{C}} P\left(\sum_{i,j} n \frac{(\hat{p}_{ij}^{(0)} - \hat{p}_{ij}^{(1)}(\pi))^2}{\hat{p}_{ij}^{(1)}(\pi)} + \mathrm{o_p}(1) \geqslant t_1, \right. \\
&\qquad \sqrt{n} \boldsymbol{\alpha}^{*\prime}(\pi) n [\mathbf{R}^0(\pi)]^{-1} \boldsymbol{\alpha}^*(\pi) + \mathrm{o_p}(1) \geqslant t_2, \\
&\qquad \left. \alpha^*_{us}(\pi) > 0, (u,s) \in \pi, \sum_{i,j} x_{ij}^{(u,s)} \hat{p}_{ij}^{(1)}(\pi) \leqslant 0, (u,s) \in \pi^c \right),
\end{aligned}
$$

where $\mathbf{R}^0(\pi)$ is as defined before. The third equality is true by (B1) for T_{01} and a result in Silvey (1959) for T_{12}. Lemma B and Lemma D, in Robertson et al. (1988, p. 71) and (3.10) imply that

$$
\lim_{n \to \infty} P(T_{01} \geqslant t_1, T_{12} \geqslant t_2) = \sum_{j=0}^{k(r-1)} a_j(\mathbf{p}^{(0)}) P(\chi^2_{k(r-1)-j} \geqslant t_1) P(\chi^2_j \geqslant t_2),
$$

which is the desired result.

Appendix D. Efficient computation of the weights associated with the asymptotic distribution of the test statistic

In this section of the Appendix we show how the estimated weights needed for obtaining the estimated asymptotic null distributions in (3.12) and (3.13) can be efficiently computed through the successive use of matrix sweeps and inversions.

Without loss of generality assume that $\pi = \{1, \ldots, d\}$, i.e. the set of constraint indices corresponding to the first d order constraints, i.e. the first d columns of $\mathbf{H}$. Partition the $\mathbf{H}$ matrix according to π as follows

$$
\mathbf{H} = [\mathbf{H}(\pi) : \mathbf{H}(\pi^c)].
$$

The corresponding partition of $\mathbf{R}^{-1} = \mathbf{H}'\mathbf{B}\mathbf{H}$ is

$$
\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{H}'(\pi)\mathbf{B}\mathbf{H}(\pi) & \mathbf{H}'(\pi)\mathbf{B}\mathbf{H}(\pi^c) \\ \mathbf{H}'(\pi^c)\mathbf{B}\mathbf{H}(\pi) & \mathbf{H}'(\pi^c)\mathbf{B}\mathbf{H}(\pi^c) \end{bmatrix}. \tag{D.1}
$$

A sweep of the matrix $\mathbf{R}^{-1}$ on its first d rows yields two matrices needed for the computation of the weights as its diagonal blocks. That is

$$\mathrm{SWEEP}(\mathbf{R}^{-1};\pi)=\begin{bmatrix}\mathbf{R}^{-1}(\pi) & \\ & \mathbf{R}^{-1}(\pi^c)-\mathbf{H}'(\pi^c)\mathbf{BH}(\pi)\mathbf{R}^{-1}(\pi)\mathbf{H}'(\pi)\mathbf{BH}(\pi^c)\end{bmatrix}.$$

Denote the two matrices shown above by

$$\begin{aligned}&\mathrm{SWEEP}_{(1)}(\mathbf{R}^{-1};\pi)=\mathbf{R}^{-1}(\pi),\\&\mathrm{SWEEP}_{(2)}(\mathbf{R}^{-1};\pi)=\mathbf{R}^{-1}(\pi^c)-\mathbf{H}'(\pi^c)\mathbf{BH}(\pi)\mathbf{R}^{-1}(\pi)\mathbf{H}'(\pi)\mathbf{BH}(\pi^c).\end{aligned}\tag{D.2}$$

Using the fact that the Sweep operator is reversible, that is SWEEP(SWEEP($\mathbf{R}^{-1}$; π); π) = $\mathbf{R}^{-1}$, we get

$$\begin{aligned}&\mathrm{SWEEP}^{-1}(\mathbf{R}^{-1};\pi)\\&\quad=(\mathrm{SWEEP}(\mathbf{R}^{-1};\pi))^{-1}\\&\quad=\begin{bmatrix}\mathbf{R}^{-1}(\pi)-\mathbf{H}'(\pi)\mathbf{BH}(\pi^c)\mathbf{R}^{-1}(\pi^c)\mathbf{H}'(\pi^c)\mathbf{BH}(\pi) & \\ & \mathbf{R}^{-1}(\pi^c)\end{bmatrix}.\end{aligned}$$

Denote the two matrices shown above by

$$\begin{aligned}&\mathrm{SWEEP}_{(3)}(\mathbf{R}^{-1};\pi)=\mathbf{R}^{-1}(\pi^c),\\&\mathrm{SWEEP}_{(4)}(\mathbf{R}^{-1};\pi)=\mathbf{R}^{-1}(\pi^c)-\mathbf{H}'(\pi^c)\mathbf{BH}(\pi)\mathbf{R}^{-1}(\pi)\mathbf{H}'(\pi)\mathbf{BH}(\pi^c).\end{aligned}\tag{D.3}$$

Similarly define the matrices in (D.2) and (D.3) for an arbitrary π with card(π) = d.

The number of πs that have cardinal d is equal to

$$m_d=\mathrm{C}_d^{(r-1)k}=\frac{[(r-1)k]!}{d![(r-1)k-d]!}.$$

Denote these by $\pi_{1,d},\ldots,\pi_{m_d,d}$. Clearly the whole set of πs with cardinal $(r-1)k-d$ is easily obtained as $\pi^c_{1,d},\ldots,\pi^c_{m_d,d}$.

From the discussion above it follows that we can compute the weights by successive sweeps and inversions using the following algorithmic scheme:

$$a_d(\mathbf{p})=\sum_{i=1}^{m_d}Q(\mathrm{SWEEP}_{(1)}(\mathbf{R}^{-1};\pi_{i,d}))Q(\mathrm{SWEEP}_{(2)}(\mathbf{R}^{-1};\pi_{i,d})),$$

$$\begin{aligned}a_{k(r-1)-d}(\mathbf{p})=&\sum_{i=1}^{m_d}Q([\mathrm{SWEEP}_{(3)}(\mathbf{R}^{-1};\pi_{i,d})]^{-1})\\&\times Q([\mathrm{SWEEP}_{(4)}(\mathbf{R}^{-1};\pi_{i,d})]^{-1}).\end{aligned}\tag{D.4}$$

For a given cardinal d, we used the SAS procedure PROC PLAN to generate all possible πs with cardinal d, i.e. all possible combinations of d rows of the $\mathbf{R}^{-1}$ matrix on which we sweep in order to evaluate the weights in (D.4). The Sweep operations and matrix inversions were done using SAS IML. Finally, we used a SAS/IML program for the calculation of the multivariate normal quadrant probabilities in (D.4). The program was written by Genz and

Bretz (contact: bretz@ifgb.uni-hannover.de) and evaluates the multivariate normal integral by applying a randomized lattice rule on a transformed integral as described by Genz (1992, 1993). It utilizes variable priorization and antithetic sampling and can compute multivariate normal probabilities for positive semi-definite covariance matrices until dimension 100.

References

Aly, E., Kochar, S., McKeague, I., 1994. Some tests for comparing cause specific hazard rates. J. Amer. Statist. Assoc. 89, 994–999.

Carriere, K.C., Kochar, S.C., 2000. Comparing sub-survival functions in a competing risks model. Lifetime Data Anal. 6, 85–97.

Dykstra, R., Kochar, S., Robertson, T., 1995. Likelihood based inference for cause specific hazard rates under order restrictions. J. Multivariate Anal. 54, 163–174.

Dykstra, R., El Barmi, H., Guffey, J., Wright, T., 1996. Nonhomogeneous Poisson processes as overhaul models. Canad. J. Statist. 24, 217–228.

El Barmi, H., Dykstra, R., 1994. Restricted multinomial maximum likelihood estimation based upon Fenchel duality. Statist. Probab. Lett. 21, 121–130.

El Barmi, H., Kochar, S., 2002. Inference for sub-survival functions under order restrictions. J. Indian Statist. Assoc. 40, 85–103.

El Barmi, H., Dykstra, R., 1995. Testing for and a against a set of linear inquality constraints in a multinomial setting. Canad. J. Statist. 23, 131–143.

Genz, A., 1992. Numerical computation of multivariate normal probabilities. J. Comput. Graph. Statist. 1, 141–149.

Genz, A., 1993. Comparison of methods for the computation of multivariate normal probabilities. Comput. Sci. Statist. 25, 400–405.

Gilbert, P.B., McKeague, I.W., Sun, Y., 2004. Tests for comparing mark-specific hazards and cumulative incidence functions. Lifetime Data Anal. 10, 5–28.

Hoel, D.G., 1972. A representation of mortality data by competing risks. Biometrics 65, 475–488.

Kochar, S.C., 1995. A review of some distribution-free tests for the equality of cause specific hazard rates. In: Koul, H.L., Desphandé, J.V. (Eds.), Analysis of Censored Data. Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 27, pp. 147–162.

Kochar, S.C., Lam, K.F., Yip, P.S.F., 2002. Generalized supremum tests for the equality of cause specific hazard rates. Lifetime Data Anal. 8, 277–288.

Kulathinal, S.B., Gasbarra, D., 2002. Testing equality of cause-specific hazard rates corresponding to m competing risks among K groups. Lifetime Data Anal. 8, 147–161.

Lam, K.F., 1998. A class of tests for the equality of k cause-specific hazard rates in a competing risks model. Biometrika 85, 179–188.

Peterson, A., 1977. Expressing the Kaplan–Meier estimator as a function of empirical sub-survival functions. J. Amer. Statist. Assoc. 72, 854–858.

Robertson, T., Wright, T., Dykstra, R., 1988. Order Restricted Statistical Inference. Wiley, New York.

Silvey, S.D., 1959. The Lagrangian multiplier test. Ann. Math. Stat. 30, 389–407.

Sun, Y., Tiwari, R.C., 1995. Comparing cause-specific hazard rates of a competing risks model with censored data. In: Koul, H.L., Desphandé, J.V., (Eds.), Analysis of Censored Data. Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 27, pp. 255–270.