TESTS BASED ON U-STATISTICS FOR TESTING EQUALITY OF COMPETING RISKS WHEN CAUSES OF FAILURE ARE MISSING[a]

Isha Dewan

Theoretical Statistics and Mathematics Unit

Indian Statistical Institute

New Delhi

[a] A joint work with **Sangita Kulathinal**

Introduction

- **Suppose a unit is subject to k competing risks. The data consists of the time to failure T and the cause of failure δ which assumes one of the values** $\{1, \ldots, k\}$.

- **The sub-distribution function** $F(j,t) = P[T \leq t, \delta = j],\ j = 1, 2, \ldots, k,$

- **The sub-survival function** $S(j,t) = P[T > t, \delta = j],\ j = 1, 2, \ldots, k.$

- **Under the assumption of absolute continuity of** $F(j,t)$, **for** $j = 1, \ldots, k$, **let** $f(j,t)$ **denote the sub-density function corresponding to risk** j**.**

- $F(t) = \sum_{j=1}^{k} F(j,t),\quad S(t) = \sum_{j=1}^{k} S(j,t)$
  $f(t) = \sum_{j=1}^{k} f(j,t).$

The cause-specific hazard rates
$\lambda(j,t) = f(j,t)/S(t),\ j = 1,\ldots,k.$

It is of interest to test whether the competing risks are equally effective.
$H_0 : F(1,t) = \ldots = F(k,t),$
$H_0 : S(1,t) = \ldots = S(k,t),$
$H_0 : \lambda(1,t) = \ldots = \lambda(k,t).$

Dewan and Deshpande (2005) reviewed the tests for testing the hypotheses of bivariate symmetry of hypothetical failure times when there are only two dependent competing risks acting in the environment.

Theorem: Under the null hypothesis of bivariate symmetry of hypothetical failure times due to two risks we have

(i) $F(1,t) = F(2,t)$ for all t,

(ii) $S(1,t) = S(2,t)$ for all t,

(iii) $\lambda(1,t) = \lambda(2,t)$ for all t,

(iv) $P[\delta = 1] = P[\delta = 2]$,

(v) T and δ are independent.

$H_0 : F(1,t) = F(2,t)$ for all t against

$H_1 : F(1,t) < F(2,t)$ for some t

$H_0 : S(1,t) = S(2,t)$ for all t against

$H_2 : S(1,t) > S(2,t)$ for some t

$H_0 : \lambda(1,t) = \lambda(2,t)$ for all t against

$H_3 : \lambda(1,t) < \lambda(2,t)$ for some t.

All these alternatives imply that risk 2 is more "effective" than risk 1 in some stochastic sense.

H_3 implies both H_1 and H_2.

These hypotheses represent different aspects of the competing risks problem. It is possible that in some cases the sub-distribution functions cross but the sub-survival functions are ordered or vice-versa.

The experimenter may have information only on failure time for the individuals. Identifying the cause for all individuals might be too expensive or just may not be feasible.

Kodel and Chen (1987) considered an example from animal bioanalysis where all causes were not available.

Lapidus et al (1994) while studyiing motorcycle fatalities observed that 40 percent of the death cerificates had no information on causes.

Several authors have considered likelihood based estimation in such situation see, for example, Dinse (1982), Dewanji(1992), Goetghebeur and Ryan (1995) , Dewanji and Sengupta (2003) , Lu and Tsiatis (2005).

Miyawaka (1984) obtained maximum likelihood estimators and minimum variance unbiased estimators of the parameters of exponenential distribution for the missing case.

Kundu and Basu (2000) discussed approximate and asymptotic properties of these estimators and obtained confidence intervals.

All above authors assume that the underlying competing risks are independent.

However, the assumption of independence of underlying competing risks may not be appropriate . Besides it can not be tested on the basis of competing risks data.

We extend some of the tests based on U-statistics for testing above hypotheses to a situtation when the causes of failure are not observed for all units under consideration and the underlying risks are not independent.

It would not be a good idea to ignore information on all failure times for which causes are failure are not observable, especially if the proportion of missing causes is large.

Let T_i $i = 1, \ldots, N$ be the failure times available on N independent units.

We consider a situation when δ_i may not be observed always $i.e.$, it may be missing for some units.

Let O_i be an indicator variable which takes value one if δ_i is observed and zero if δ_i is missing.

We assume that δ_i are missing at random and hence O_i is independent of (T_i, δ_i).
Similar assumptions are made by Rubin (1976) and Lu and Tsiatis (2005).

Let us assume that the causes of failure are available for only n units out of N. Note that n is a random number.

Tests for H_0 against H_1

$H_0 : F(1,t) = F(2,t) = F(t)/2,$ for all t

against the alternative

$H_1 : F(1,t) < F(2,t).$

A measure of deviation from the null hypothesis is $F(2,t) - F(1,t)$ which is nonnegative under H_1.

Consider $\int_0^\infty [F(2,t) - F(1,t)]dF(t)$
$= 2P[T_1 \leq T_2, \delta_1 = 2] - \frac{1}{2}.$

When $(T_i, \delta_i), i = 1, \ldots, N$ are available, a U-statistic (U_F) for testing H_0 against H_1 is based on the kernel, which is an estimator of the above measure without the constant term, defined as follows

$$\begin{aligned} \phi_F(T_i,\delta_i,T_j,\delta_j) &= 1 \text{ if } T_i < T_j, \delta_i = 2 \\ &\quad \text{or, } T_j < T_i, \delta_j = 2, \\ &= 0 \text{ otherwise.} \end{aligned}$$

Let U_F be the U-statistic corresponding to the kernel $\phi_F(.)$

$U_F = \frac{1}{\binom{N}{2}} \sum_{1\le i<j\le N} \phi(T_i,\delta_i,T_j,\delta_j)$.

Then $EU_F = 2\int_0^\infty F(2,t)dF(t)$,
$E(U_F \mid H_0) = 1/2$, $E(U_F \mid H_1) > 1/2$,
$var(U_F \mid H_0) = 1/3$.

In case if δ_i for some i are missing then the above kernel can not be defined for each pair (i,j).

Following table shows the situations when the kernel can not be defined (m indicates missing δ and ? indicates that the kernel can not be defined).

(δ_i, δ_j)	(1,1)	(1,2)	(1,m)	(2,1)	(2,2)	(2,m)
$T_i > T_j$	0	1	?	0	1	?
$T_i \leq T_j$	0	0	0	1	1	1

(δ_i, δ_j)	(m,1)	(m,2)	(m,m)
$T_i > T_j$	0	1	?
$T_i \leq T_j$	?	?	?

Out of 18 combinations of a pair of observations, the kernel can not be defined for 6 combinations when either δ_i or δ_j or both are missing.

In such cases, the observations neither support or negate the null hypothesis but the information on the ordering of failure times is fully available.

Hence, in order to retrieve the best possible information we assign weight $1/2$ for these 6 combinations.

$$
\begin{aligned}
& \phi_{FM}(T_i, \delta_i, O_i, T_j, \delta_j, O_j) \\
= \; & 1 \text{ if } T_i < T_j, \delta_i = 2, O_i = 1 \\
& \text{ or, } T_j < T_i, \delta_j = 2, O_j = 1, \\
= \; & 1/2 \text{ if } T_i < T_j, O_i = 0 \\
& \text{ or, } T_j < T_i, O_j = 0, \\
= \; & 0 \text{ otherwise.}
\end{aligned}
$$

The corresponding U-statistic is
$U_{FM} = \frac{1}{\binom{N}{2}} \sum_{1 \le i < j \le N} \phi(T_i, \delta_i, O_i, T_j, \delta_j, O_j)$.

$EU_{FM} = 2p \int_0^\infty F(2,t)dF(t) + (1-p)/2$
$= pE(U_F) + (1-p)/2$, where $p = pr(O_i = 1)$ is the probability that δ is observed.

Under H_0, $E(U_{FM}) = 1/2$
and under H_1, $E(U_{FM}) > 1/2$.

A straightforward computation shows that under H_0

$$\lim_{n \to \infty} var(\sqrt{N} U_{FM}) = 4[(p+3)/12 - 1/4] = p/3.$$

If $p = 1$ that is there are no missing causes then the variance is equal to $1/3$ which is the asymptotic variance of $\sqrt{N}U_F$.

As p decreases, variance of $\sqrt{N}U_{FM}$ linearly decreases and is 0 if causes are missing for all units.

In practice, p is generally unknown and hence, the variance need to be estimated by replacing p by its empirical estimator, $\hat{p} = n/N$.

Using the results from Serfling (1980) it follows that under H_0 $\sqrt{N}(U_{FM} - 1/2)$ converges in distribution to normal random variable with mean zero and variance $p/3$ as $N \to \infty$.

Hence we reject H_0 if $\sqrt{3N}(U_{FM} - 1/2)/\sqrt{\hat{p}} > Z_\alpha$ where Z_α is the upper α point of the null distribution of standard normal distribution.

Exact Null Distribution

Let R_i be the rank of T_i among $T_1, T_2, \ldots, T_N$.

$$\binom{N}{2} U_{FM} = 1/2 \sum_{i=1}^{N} [N - i + 1](1 - O_{(i)})$$
$$+ \sum_{i=1}^{N} [N - i + 1] O_{(i)} I(\delta_{(i)} = 2),$$

where $O_{(i)}$ and $\delta_{(i)}$ denote the observations corresponding to $T_{(i)}$, ith ordered T.

If all causes of failure are observed then
$$\binom{N}{2} U_{FM} = \sum_{i=1}^{N} [N - R_i + 1] I(\delta_i = 2).$$

The joint distribution of O_i and δ_i is specified by
$$pr(O_i = 0, \delta_i = 1) = (1 - p)\theta,$$

$$pr(O_i = 0, \delta_i = 2) = (1 - p)(1 - \theta),$$

$$pr(O_i = 1, \delta_i = 1) = p\theta,$$

$$pr(O_i = 1, \delta_i = 2) = p(1-\theta),$$

where $\theta = pr(\delta_i = 1) = 1 - pr(\delta_i = 2)$.

Under $H_0,\ r = 0, 1; s = 1, 2$,
$pr(O_i = r, \delta_i = s) = pr(O_{(i)} = r, \delta_{(i)} = s)$.

The moment generating function of $\binom{N}{2} U_{FM}$ under H_0 is

$$M(t) = \prod_{i=1}^{N} [p\theta + p(1-\theta)\exp\{t(N-i+1)\}$$
$$+(1-p)\exp\{t(N-i+1)/2\}].$$

Under H_0, $\theta = 1/2$ and the moment generating function depends on the unknown p.

Hence, the statistic is not distribution free even under the null hypothesis.

Note that p is a nuisance parameter in the sense that it is extraneous to the joint distribution of (T, δ).

Large positive values of the statistic support the alternative hypothesis.

Tests for H_0 against H_3

$$H_0 : \lambda(1,t) = \lambda(2,t),\ \forall t$$

against the alternative hypothesis
$H_1 : \lambda(1,t) < \lambda(2,t),$ for some t.

Kochar (1995) and Dewan and Deshpande (2005) for a review of various test procedures for H_0 against H_3 when data on causes of failures are fully observed.

$$\Psi(t) = F(1,t) - F(2,t)$$
$$= \int_0^t S(u)[\lambda(1,u) - \lambda(2,u)]du$$

is negative and non-increasing in t iff H_3 holds.

$\Delta = \int_{0<x<y<\infty} [\Psi(x) - \Psi(y)] dF(x) dF(y).$

Simplify Δ as

$$\begin{aligned} \Delta &= 2 \int_{0<x<y<\infty} [F(1,x) - F(1,y)] dF(x) dF(y) \\ &\quad - \int_{0<x<y<\infty} [F(x) - F(y)] dF(x) dF(y) \\ &= 4pr(T_1 < T_2 < T_3, \delta_1 = 1) \\ &\quad -2pr(T_1 < T_2, \delta_1 = 1) + 1/6. \end{aligned}$$

U_{1H} denotes the U-statistic estimator of $pr(T_1 < T_2 < T_3, \delta_1 = 1)$ and U_{2H} denotes the U-statistic estimator of $pr(T_1 < T_2, \delta_1 = 1)$.

These statistics are based on the following kernels, respectively.

$$
\begin{aligned}
\phi_{1H}(T_i, \delta_i, T_j, \delta_j, T_k, \delta_k) &= 1 \text{ if } T_i < T_j < T_k, \delta_i = 1 \\
&\quad \text{or } T_i < T_k < T_j, \delta_i = 1 \\
&\quad \text{or } T_j < T_i < T_k, \delta_j = 1 \\
&\quad \text{or } T_j < T_k < T_i, \delta_j = 1 \\
&\quad \text{or } T_k < T_i < T_j, \delta_k = 1 \\
&\quad \text{or } T_k < T_j < T_i, \delta_k = 1 \\
&= 0 \text{ otherwise,}
\end{aligned}
$$

$$
\begin{aligned}
\phi_{2H}(T_i, \delta_i, T_j, \delta_j) &= 1 \text{ if } T_i < T_j, \delta_i = 1 \\
&\quad \text{or } T_j < T_i, \delta_j = 1, \\
&\quad 0 \text{ otherwise.}
\end{aligned}
$$

$$U_{1H} = \frac{1}{\binom{N}{3}} \sum_{1 \le i < j < k \le N} \phi_{1H}(T_i, \delta_i, T_j, \delta_j, T_k, \delta_k)$$

$$U_{2H} = \frac{1}{\binom{N}{2}} \sum_{1 \le i < j \le N} \phi_{2H}(T_i, \delta_i, T_j, \delta_j).$$

$$E(U_{1H}) = 6pr(T_1 < T_2 < T_3, \delta_1 = 1)$$

and $E(U_{2H}) = 2pr(T_1 < T_2, \delta_1 = 1)$

Consider the statistic
$U_H = 2/3\, U_{1H} - U_{2H},$
- a linear combination of U_{1H} and U_{2H} .
$E(U_H) = 4pr(T_1 < T_2 < T_3, \delta_1 = 1) - 2pr(T_1 < T_2, \delta_1 = 1)$
$= \Delta - 1/6.$
Under H_0, $E(U_H) = -1/6$, and under H_3 , $E(U_H) > -1/6$.

We extend both the kernels to missing data case as follows.

$$
\begin{aligned}
&\phi_{1HM}(T_i, \delta_i, O_i, T_j, \delta_j, O_j, T_k, \delta_k, O_k) \\
&= 1 \text{ if } T_i < T_j < T_k, \delta_i = 1, O_i = 1 \\
&\quad \text{or } T_i < T_k < T_j, \delta_i = 1, O_i = 1 \\
&\quad \text{or } T_j < T_i < T_k, \delta_j = 1, O_j = 1 \\
&\quad \text{or } T_j < T_k < T_i, \delta_j = 1, O_j = 1 \\
&\quad \text{or } T_k < T_i < T_j, \delta_k = 1, O_k = 1 \\
&\quad \text{or } T_k < T_j < T_i, \delta_k = 1, O_k = 1 \\
&1/2 \text{ if } T_i < T_j < T_k, O_i = 0 \\
&\quad \text{or } T_i < T_k < T_j, O_i = 0 \\
&\quad \text{or } T_j < T_i < T_k, O_j = 0 \\
&\quad \text{or } T_j < T_k < T_i, O_j = 0 \\
&\quad \text{or } T_k < T_i < T_j, O_k = 0 \\
&\quad \text{or } T_k < T_j < T_i, O_k = 0 \\
&0 \text{ otherwise,}
\end{aligned}
$$

$$
\begin{aligned}
&\phi_{2HM}(T_i, \delta_i, O_i, T_j, \delta_j, O_j) \\
= \;& 1 \;\text{ if }\; T_i < T_j, \delta_i = 1, O_i = 1 \\
& \quad \text{or }\; T_j < T_i, \delta_j = 1, O_j = 1 \\
& 1/2 \;\text{ if }\; T_i < T_j, O_i = 0 \\
& \quad \text{or, }\; T_j < T_i, O_j = 0 \\
& 0 \;\text{ otherwise.}
\end{aligned}
$$

$U_{HM} = 2/3\, U_{1FM} - U_{2FM}$,
where U_{1HM} and U_{2HM} are U-statistics defined using the corresponding kernels.

$$
\begin{aligned}
& E(\phi_{1HM}(T_i, \delta_i, O_i, T_j, \delta_j, O_j, T_k, \delta_k, O_k)) \\
= \; & 6p\, pr(T_1 < T_2 < T_3, \delta_1 = 1) + (1-p)/2,
\end{aligned}
$$

$$
\begin{aligned}
& E(\phi_{2HM}(T_i, \delta_i, O_i, T_j, \delta_j, O_j)) \\
= \; & 2p\, pr(T_1 < T_2, \delta_1 = 1) + (1-p)/2.
\end{aligned}
$$

$$
\begin{aligned}
E(U_{HM}) &= 4p\, pr(T_1 < T_2 < T_3, \delta_1 = 1) + (1-p)/3 \\
&\quad -2p\, pr(T_1 < T_2, \delta_1 = 1) - (1-p)/2 \\
&= p(\Delta - 1/6) - (1-p)/6 \\
&= pE(U_H) - (1-p)/6.
\end{aligned}
$$

Large values of the statistic are significant.

Under H_0, the asymptotic variances and covariance are

$$\begin{aligned} var(\sqrt{N}\, U_{1HM}) &= 9p/20, \\ var(\sqrt{N}\, U_{2HM}) &= 4p/12 = p/3, \\ N\ cov(U_{1HM}, U_{2HM}) &= 6p/16 = 3p/8. \end{aligned}$$

Hence, the asymptotic variance of
$var(\sqrt{N}\, U_{HM}) = 4/9 * 9p/20 + p/3 - 4/3 * 3p/8$
$= p/30.$

Under H_0, $\sqrt{N}(U_{1HM}) - EU_{1HM}))$ converges in distribution to $N(0, 9p/20)$
and $\sqrt{N}(U_{2HM}) - EU_{2HM}))$ converges in distribution to $N(0, p/3)$,

so that $\sqrt{N}(U_{HM}) - EU_{HM}))$ converges in distribution to $N(0, p/30)$.

Thus for large sample sizes we can use critical points from standard normal distribution.
Ofcourse one would need to replace p by its consistent estimator.

$$\binom{N}{3}U_{1HM} = \sum_{i=1}^{N}\binom{N-R_i}{2}O_iI(\delta_i = 1)$$
$$+1/2\sum_{i=1}^{N}\binom{N-R_i}{2}(1-O_i),$$

$$\binom{N}{2}U_{2HM} = \sum_{i=1}^{N}(N-R_i)O_iI(\delta_i = 1)$$
$$+1/2\sum_{i=1}^{N}(N-R_i)(1-O_i)$$

Tests for H_0 against H_2

$H_0 : S(1,t) = S(2,t), \ \forall \ t$
against $H_2 : S(1,t) > S(2,t),$ for some t.

Again a measure of deviation from the null hypothesis is $S(1,t) - S(2,t)$ which is zero under H_0 and nonnegative under H_2.

Kochar (1995) considered tests based on U-statistics and Carriere and Kochar (2000) considered Kolmogrov-Smirnov type types based on maximum differences between sub-survival functions.

Consider the following distance measure
$\int_0^\infty [S(1,t) - S(2,t)] dF(t)$
$= 2P[T_1 > T_2, \delta_1 = 1] - \frac{1}{2}$.

The kernel defined below was used to test H_0 against H_2

$$\begin{aligned}\phi_S(T_i, \delta_i, T_j, \delta_j) &= 1 \text{ if } T_i > T_j, \delta_i = 1 \\ &\quad \text{or, } T_j > T_i, \delta_j = 1, \\ &= 0 \text{ otherwise}\end{aligned}$$

$$E(\phi_S(T_i, \delta_i, T_j, \delta_j)) = 2P[T_1 > T_2, \delta_1 = 1].$$

However, the above statistic is not completely defined if the cause of failure δ is missing for some units.

The table below shows the situations when the kernel can not be defined (m indicates missing δ and ? indicates that the kernel can not be defined).

(δ_i, δ_j)	(1,1)	(1,2)	(1,m)	(2,1)	(2,2)	(2,m)
$T_i > T_j$	1	1	1	0	0	0
$T_i \leq T_j$	1	0	?	1	0	?

(δ_i, δ_j)	(m,1)	(m,2)	(m,m)
$T_i > T_j$	?	?	?
$T_i \leq T_j$	1	0	?

The Statistic

We replace ? by $1/2$, so that the modified kernel is

$$\begin{aligned} & \phi_{SM}(T_i, \delta_i, O_i, T_j, \delta_j, O_j) \\ = \; & 1 \text{ if } T_i > T_j, \delta_i = 1, O_i = 1 \\ & \quad \text{or, } T_j > T_i, \delta_j = 1, O_j = 1, \\ = \; & 1/2 \text{ if } T_i > T_j, O_i = 0 \\ & \quad \text{or, } T_j > T_i, O_j = 0, \\ = \; & 0 \text{ otherwise.} \end{aligned}$$

The U-statistic is

$$U_{SM} = \frac{1}{\binom{N}{2}} \sum_{1 \le i < j \le N} \phi(T_i, \delta_i, O_i, T_j, \delta_j, O_j).$$

$E(U_{SM}) = 2p \int_0^\infty S(1,t)dF(t) + (1-p)/2$
$= pE(U_S) + (1-p)/2,$
where U_S is the U-statistic corresponding to the kernel $\phi_S(T_i, \delta_i, T_j, \delta_j)$.

Expectation is $1/2$ under the null hypothesis and is greater than $1/2$ under H_2.

Under H_0 the asymptotic variance of $\sqrt{N}U_{SM} = p/3$.

The limiting distribution of $\sqrt{N}(U_{SM} - 1/2)$ is normal with mean zero and variance $p/3$.

Large positive values of the statistic support the alternative hypothesis.

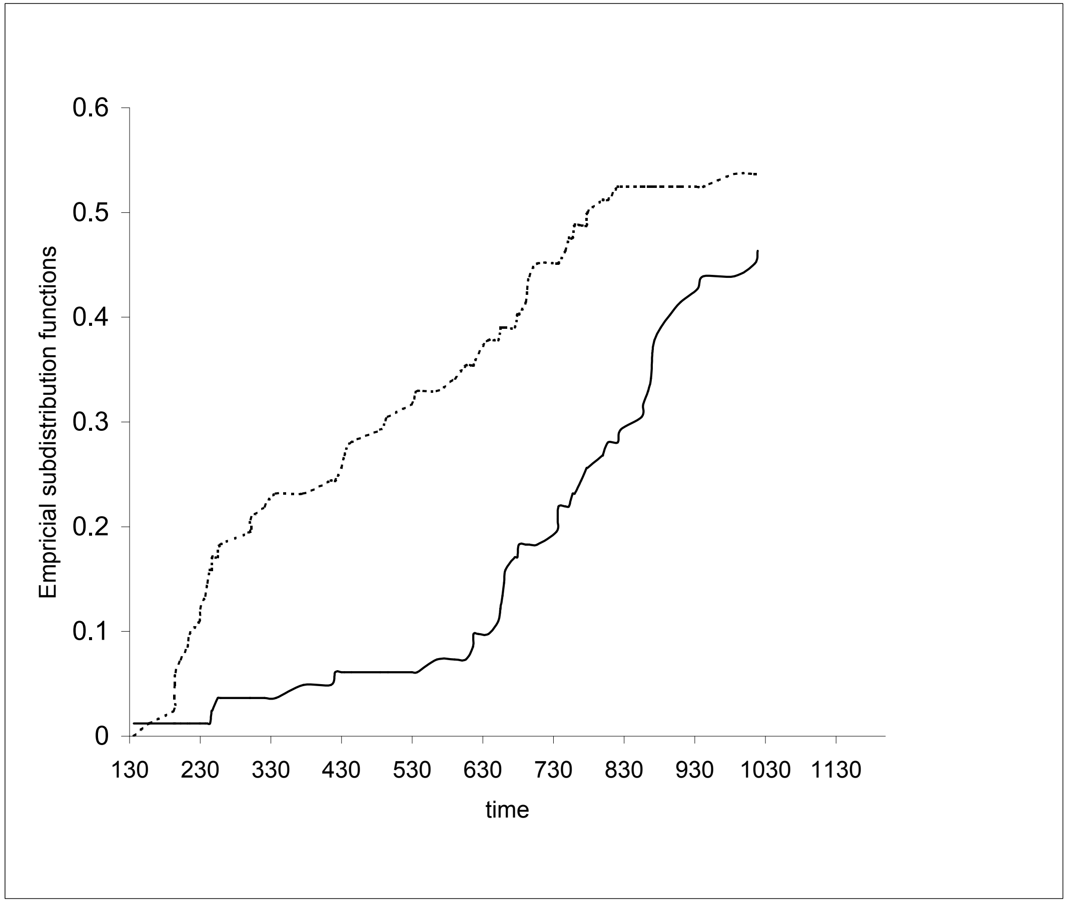
Illustration

We apply the proposed tests to the data from Hoel (1972).

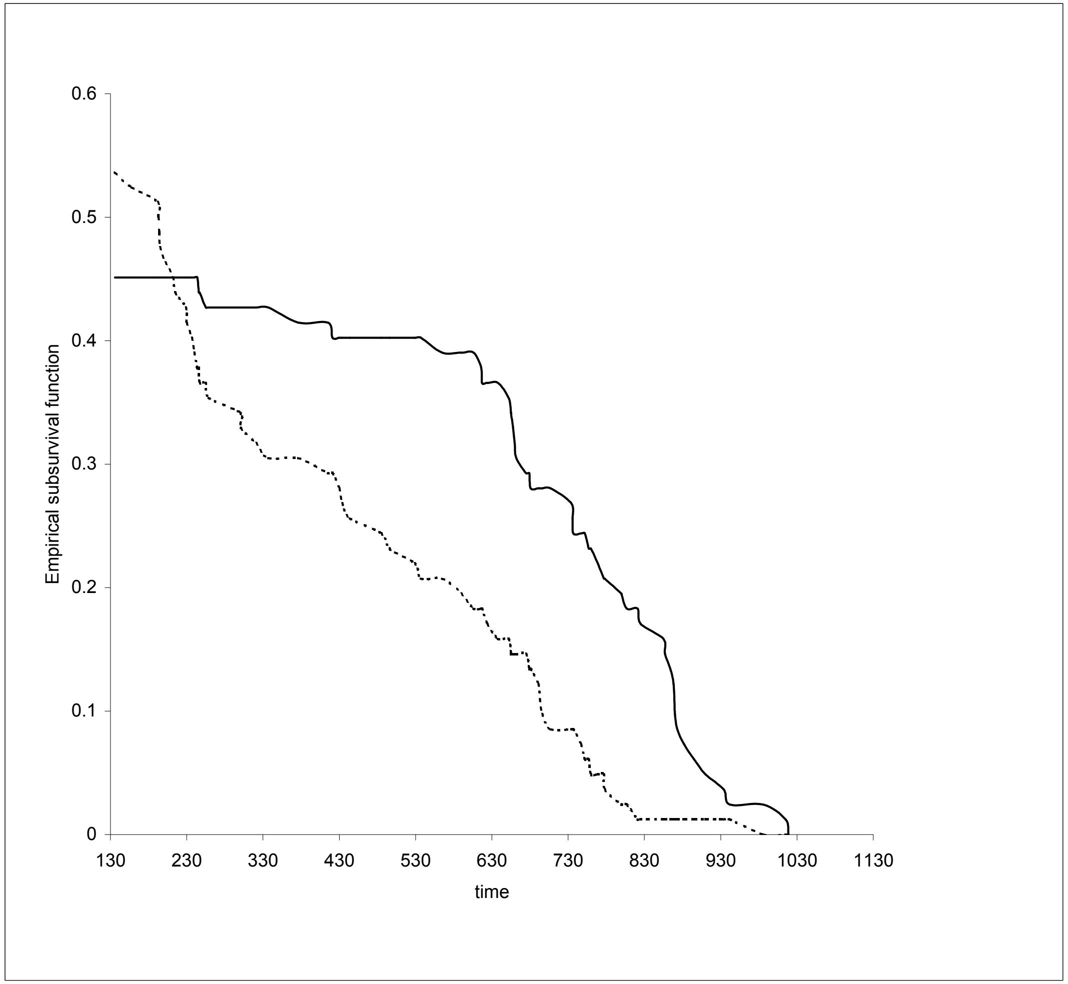
The data were obtained from a laboratory experiment on two groups of RFM strain male mice which had received a radiation dose of 300r at an age of 5-6 weeks and which lived in conventional laboratory environment and in germ-free environment.

Here we analyse data on $N = 82$ mice which lived in germ-free environment and consider two causes of death, cancer (thymic lymphoma and reticulum cell sarcoma combined as risk 2) and other causes (risk 1) as two competing risks.

- Figure 1 - Empirical sub-distribution function for Hoel's data



- Figure 2 - Empirical sub-survival function for Hoel's data



We artifically created missing data for the causes with p equal to $0.95, 0.90, 0.80, 0.70, 0.60, 0.50$.

The performance of the proposed tests for these values of p is compared by evaluating the test statistics in three cases;

(a) no missing data (sample size = N, $p = 1$),

(b) adjusting for missing causes by using U_{*M} test statistic (sample size = N, p known) and

(c) by excluding observations with missing causes and reducing the sample size accordigly (sample size = n, $p = 1$), where $*$ is either F, S or H.

In the case (c) the sample size n is random.

However, we calculate the statistics as if n is fixed.

Table 1: Comparison of test statistics for hypothses $H_0 : F(1,t) = F(2,t)$ for various proportions of missing data

((b) = $\sqrt{N}(U_{FM} - 0.5)/\sqrt{(p/3)}$ and (c) = $\sqrt{n}(U_F - 0.5)/\sqrt{(1/3)}$)

p	n	(b)	(c)
1	82	2.925	2.925
0.95	79	3.207	3.170
0.9	72	2.907	3.041
0.8	62	2.323	2.383
0.7	55	2.249	2.296
0.6	52	2.890	2.693
0.5	38	1.058	1.222

$H_0 : F(1,t) = F(2,t)$ is rejected against $H_1 : F(1,t) < F(2,t)$ at 5% level of significance if the value of the normalised test statistics is larger than 1.44.

Table 2: Comparison of test statistics for hypothses $H_0 : S(1,t) = S(2,t)$ for various proportions of missing data

((b) = $\sqrt{N}(U_{SM} - 0.5)/\sqrt{(p/3)}$ and (c) = $\sqrt{n}(U_S - 0.5)/\sqrt{(1/3)}$)

p	n	(b)	(c)
1	82	1.792	1.792
0.95	79	1.456	1.431
0.9	72	1.309	1.425
0.8	62	1.483	1.510
0.7	55	1.123	1.146
0.6	52	1.920	1.742
0.5	38	1.619	1.814

In Table 2, $H_0 : S(1,t) = S(2,t)$ is rejected against $H_2 : S(1,t) > S(2,t)$ at 5% level of significance if the value of the normalised test statistics is larger than 1.44.

Table 3: Comparison of test statistics for hypothses $H_0 : \lambda(1,t) = \lambda(2,t)$ for various proportions of missing data

((b) = $\sqrt{N}(U_{HM} + 1/6)/\sqrt{(p/30)}$ and (c) = $\sqrt{n}(U_H + 1/6)/\sqrt{(1/30)}$)

p	n	(b)	(c)
1	82	0.855	0.855
0.95	79	1.221	1.299
0.9	72	1.286	1.355
0.8	62	0.672	0.643
0.7	55	0.522	0.659
0.6	52	0.753	1.00
0.5	38	-0.237	-0.229

$H_0 : \lambda(1,t) = \lambda(2,t)$ is rejected against $H_3 : \lambda(1,t) < \lambda(2,t)$ at 5% level of significance if the value of the normalised test statistics is larger than 1.44.

Discussion

We feel that newly proposed tests would perform better than the existing tests with N replaced by n and $p = 1$.

One should note that in the problem with missing data n is random. Such tests are conditional tests, conditioned on $N = n$.

The distribution theory of unconditional tests would be complicated.

It would not be desirable to ignore information on $N - n$ failure times for which causes are not known as they to give some idea about departure from relevant null hypotheses.

We are looking at Kaplan-Meier type estimators for the sub-distribution function and the sub-survival function for missing data and also Kolmogrov-Smirnov type tests based on maximum differences between sub-distribution functions, sub-survival functions when the information on cause of failure is partly avavilable.

References1

- Aly, Emad-Eldin A. A., Kochar, S.C. and McKeague, I. W. (1994). Some tests for comparing cumulative incidence functions and cause-specific hazard rates. *J. Amer. Statist. Assoc.* , **89** , 994-999.
- Carriere, K. C. and Kochar, S.C. (2000). Comparing sub-survival functions in a competing risks model. *Lifetime Data Anal.* , **6** , 85-97.
- Dewan, I., Deshpande, J.V. and Kulathinal, S.B. (2004). On testing dependence between time to failure and cause of failure via conditional probabilities. *Scandanavian Journal of Statistics*, **31**, 79-92
- Dewan, I.and Deshpande, J.V. (2005). Tests for some statistical hypotheses for dependent competing risks - A review. To appear.
- Dewanji, A. (1992). A note on a test for competing risks with general missing pattern in failure types. *Biometrics*, **79**, 855-857.

References 2

- Dewanji, A.and Sengupta, D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics*, **59** , 1063-1070.
- Dinse , G.E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. *J. Amer. Statist. Assoc.* , **81** , 328-336.
- Goetghebeur, E. and Ryan, L. (1990). A modified log rank test for competing risks with missing failure type. *Biometrika* , **77**, 207–211.
- Goetghebeur,E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika* , **82** , 821–833.
- Hoel, D. G. (1972). A representation of mortality data by competing risks. *Biometrics*, **28**, 475-488.

References 3

- Kochar, S.C. (1995). A review of some distribution-free tests for the equality of cause specific hazard rates. *Analysis of Censored Data* , ed Koul, H.L. and Deshpande, J.V., IMS, Hayward, 147-162.
- Kodell, R.K. and Chen, J.J. (1987). Handling cause of death in equivocal cases using teh EM algorithm. *Comm. Stats. - Theory and Methods* , **16**, 2565-2603.
- Kundu, D. and Basu, S. (2000). Analysis of incomplete data in presence of competing risks. *J. Statist. Plann. Inference* , **87** , 221–239.
- Lapidus , G. Braddock, M. , Schwartz , R. Banco, L. and Jacobs, L. (1994). Accuracy of fatal motorcycle injury reporting on death certificates. *Accident Anal. Prevention* , **26**, 535-542.
- Lee, A. J. (1990). *U-statistics. Theory and practice.*. Marcel Dekker, New York,

References 4

- Lu, K. and Tsiatis, A. (2005). Comparison between two partial likelihood approaches for the competing risks model with missing cause of failure. *Lifetime Data Anal.*, **11** , 29–40.
- Rubin , D.R. (1976). Inference and missing data. *Biometrika* , *63*, 581-592.
- Serfling , R.J. (1980). *Approximation theorems of mathematical statistics.* John Wiley , New York.