Nonparametric Predictive Comparison of Two Groups of Lifetime Data

F.P.A. COOLEN
University of Durham, United Kingdom

K.J. YAN
University of Durham, United Kingdom

Abstract

We present the application of a recently introduced nonparametric predictive inferential method to compare two groups of data, consisting of observed event times and right-censoring times. Comparison is based on imprecise probabilities concerning one future observation per group.

Keywords

censored data, exchangeability, nonparametrics, prediction, survival analysis

1 Introduction

We apply a recently introduced method for statistical inference, called 'nonparametric predictive inference' (NPI) [1, 6], to the problem of comparing two groups of data, or, if one prefers to use such terminology, two underlying populations, where the data include right-censored observations. This generalizes the results presented by Coolen [3], who did not allow censoring. Right-censoring typically occurs in study of event times, e.g. survival times of patients in medical applications, or periods without failures of technical systems in reliability engineering, where a right-censoring at a time t just implies that the event of interest has not yet happened before or at time t. Throughout, we assume that no further information is available about the random quantities corresponding to right-censored observations, an assumption often called 'noninformative censoring' [6, 11, 13]. We also assume that the two populations compared are independent, in the sense that any information about the random quantities from one population does not influence our inferences on random quantities from the other population.

The method of statistical inference used here is based on quite minimal modelling assumptions, and is directly in terms of random quantities representing future observations. We assume that either a well-specified event happens, at a

particular time, to each item for which, or individual for who, we have an observation, or that a time is reported at which such an event has not yet occurred. All data are referred to as 'observation (time)', if it is a time at which the event of interest actually occurred we call it 'event (time)', else '(right-)censoring (time)'. Speaking in terms of 'time', we restrict attention to non-negative random quantities, so to random quantities and observations on the time-axis $[0,\infty)$. However, the method presented is more widely applicable, as only a finite partition of (part of) the real line is required.

In Section 2, the basics of nonparametric predictive inference are briefly summarized. Section 3 presents the main result on predictive comparison of two groups of lifetime data, which is illustrated, and briefly compared with an alternative nonparametric method, via two examples in Section 4. For ease of notation, we assume that there are no ties of any kind in the data, so no two observations are equal. In Section 5, we briefly discuss how the method can be adapted for dealing with tied observations, and we add a few concluding remarks about the presented method and results, including some attention to when this method might be used.

2 Nonparametric predictive inference

In this section, we summarize NPI for data including right-censored observations, as recently presented by Coolen and Yan [6], to which we refer for the theoretical justification and further detailed discussion of this method.

Let a single group of data consist of n observations, of which u are event times, $0 < t_1 < \ldots < t_u$, and $v = n-u$ right-censoring times, $0 < c_1 < \ldots < c_v$. Let $t_0 = 0$ and $t_{u+1} = \infty$, and let the right-censoring times in (t_i, t_{i+1}) be $c_1^i < \ldots < c_{l_i}^i$. We assume that there are no ties among the data, the method is easily adapted for ties [6]. Let $\tilde{n}_t$ be the number of items with observation time greater than or equal to t. We call this the number of items 'at risk just prior to time t', at an observation time the corresponding item is included in $\tilde{n}_t$.

Based on such data, Coolen and Yan [6] introduce, and justify, the assumption 'right-censoring $A_{(n)}$' (rc-$A_{(n)}$) for NPI, for the random quantity X_{n+1} representing the lifetime of a future item, or the survival time of a future individual. Right-censoring $A_{(n)}$ generalizes Hill's $A_{(n)}$ [7], which underlies NPI if the data do not include right-censored observations [1, 3]. Description of rc-$A_{(n)}$ requires notation for partial specification of probability distributions, called 'M-function'.

Definition 1 (M-function) *[6]*
A partial specification of a probability distribution for a real-valued random quantity X can be provided via probability masses assigned to intervals, without any further restriction on the spread of the probability mass within each interval. A probability mass assigned, in such a way, to an interval (a,b), is denoted by $M_X(a,b)$, and referred to as M-function value for X on (a,b).

Clearly, all M-function values for X on all intervals should sum up to one, and each M-function value should be in $[0,1]$.

Definition 2 (rc-$A_{(n)}$) *[6]*
The assumption 'right-censoring $A_{(n)}$' (rc-$A_{(n)}$) is that the probability distribution for a nonnegative random quantity X_{n+1}, based on u event times and v right-censoring times, as described above, is partially specified by ($i = 0,\ldots,u$; $k = 1,\ldots,l_i$)

$$M_{X_{n+1}}(t_i,\, t_{i+1}) = \frac{1}{n+1} \prod_{\{r:\, c_r < t_i\}} \frac{\tilde{n}_{c_r}+1}{\tilde{n}_{c_r}},$$

$$M_{X_{n+1}}(c_k^i,\, t_{i+1}) = \frac{1}{(n+1)\tilde{n}_{c_k^i}} \prod_{\{r:\, c_r < c_k^i\}} \frac{\tilde{n}_{c_r}+1}{\tilde{n}_{c_r}}.$$

The product terms are defined as one if the product is taken over an empty set. The M-function values for X_{n+1} on other intervals are zero. This implicitly assumes non-informative censoring, as a post-data assumption related to exchangeability of all items known to be at risk at any time t, see Coolen and Yan [6], who also justify rc-$A_{(n)}$. We illustrate the M-function values in rc-$A_{(n)}$ via an example, followed by a brief explanation of the key ideas behind rc-$A_{(n)}$.

Example 1
Table 1 gives the data for group A which are part of Example 2 in Section 4, where the data are introduced in more detail. For this group, there are 10 observed event times and 6 right-censoring times. Table 1 also presents the M-function values, with corresponding intervals, according to rc-$A_{(n)}$ for these data.

These M-function values sum up to one (subject to a minor rounding effect), and illustrate the effects of right-censoring. Notice, for example, that there is some probability mass defined on each interval from a right-censoring time to the next observed event time, and that a right-censored observation also leads to larger M-function values between two later observed event times.

This assumption rc-$A_{(n)}$ is generalizing Hill's assumption $A_{(n)}$ [7], the idea is roughly as follows. If $n+1$ real-valued random quantities are exchangeable, and we assume that ties occur with probability zero, then the $n+1$-st of these random quantities has equal probability $1/(n+1)$ to fall in each of the intervals that form the partition created by the values of the other n random quantities, *before* any of these random quantities are actually observed. Hill [7] proposed this same property as a *posterior* predictive distribution, calling it $A_{(n)}$, and later he [8, 9] discussed further properties of this assumption and its use as an inferential procedure, and presented a prior process that leads to $A_{(n)}$ in the Bayesian framework (under finite additivity). Generally speaking, use of $A_{(n)}$ makes sense in case of very vague prior information, or indeed if one explicitly wishes not to use any

Table 1: Cervical cancer example ($>t$: right-censoring at t)

data		value
	$M(0,90)$	0.05882
90	$M(90,142)$	0.05882
142	$M(142,150)$	0.05882
150	$M(150,269)$	0.05882
269	$M(269,291)$	0.05882
291	$M(291,680)$	0.05882
$>$468	$M(468,680)$	0.00535
680	$M(680,837)$	0.06417
837	$M(837,1037)$	0.06417
$>$890	$M(890,1037)$	0.00802
1037	$M(1037,1297)$	0.07219
$>$1090	$M(1090,1297)$	0.01203
$>$1113	$M(1113,1297)$	0.01684
$>$1153	$M(1153,1297)$	0.02527
1297	$M(1297,1429)$	0.12634
1429	$M(1429,\infty)$	0.12634
$>$1577	$M(1577,\infty)$	0.12634

such prior information. Our generalization adopts the same idea for the situation of right-censored data, using the extra assumption that a right-censored item, at the moment the censoring takes place, had an exchangeable residual time till event with all those items for which the event had not yet taken place, and which had not been censored previously. This exchangeability at time of censoring is indeed a proper form of 'noninformative censoring', and the probabilities as specified by rc-$A_{(n)}$, via M-function values, for a single future observation are then derived via conditioning on possible values for the right-censored items. Further details of the derivation and justification of rc-$A_{(n)}$ are given by Yan [16] and Coolen and Yan [6].

Berliner and Hill [2] also presented the use of $A_{(n)}$ for right-censored data, but instead of adding an assumption to deal with the exact censoring information, they replaced each censored observation by just survival past the largest observed event time smaller than the censoring time, in which case no assumptions need to be added to $A_{(n)}$. This implies that at observed event times, our method coincides with the Berliner-Hill method, but these two methods differ in between event times if there are censoring times. In addition, Berliner and Hill assumed that the probability mass per interval is uniformly distributed (except for the last interval if there is no finite right-end point), whereas we use imprecise probabilities, as we discuss next.

It should be mentioned that, of course, imprecise probabilities have been used before for situations where not all data are complete, in the sense that not each event of interest has actually been observed. For example, Manski [12] considers the logical bounds on conditional probabilities based on censored samples alone. This would relate to our approach if we had not added any further assumption about the right-censored data, the novelty of rc-$A_{(n)}$ is the extra exchangeability-related assumption about the residual time till event for each censored observation, which has the effect of keeping imprecision relatively small, which is particularly useful if there are relatively many censored observations in the data set.

The partial specification of the probability distribution of X_{n+1}, via M-function values as specified by rc-$A_{(n)}$, enables NPI if the problems considered can be formulated in terms of a future observation X_{n+1}. However, for many problems of interest, the M-function values only imply bounds for predictive probabilities, where optimal bounds are imprecise probabilities [15].

As a consequence of the M-function values defined in rc-$A_{(n)}$, the events $\{X_{n+1} \in (t_i,\ t_{i+1})\}$, for $i = 0, \ldots, u$, have precise probabilities [6]

$$P(X_{n+1} \in (t_i,\ t_{i+1}))\ =\ M_{X_{n+1}}(t_i,\ t_{i+1}) + \sum_{k=1}^{l_i} M_{X_{n+1}}(c_k^i,\ t_{i+1}).$$

3 Comparing two groups of lifetime data

For the comparison of two groups of lifetime data we use the notation as introduced above, but consistently add an index a or b, corresponding to the groups which we call A and B. For example, for group A we have n_a observations, consisting of the event times $0 < t_{a,1} < \ldots < t_{a,u_a}$ and right-censoring times $0 < c_{a,1} < \ldots < c_{a,v_a}$, and the right-censoring times in the interval $(t_{a,i}, t_{a,i+1})$ are denoted by $c^i_{a,1} < \ldots < c^i_{a,l_{a,i}}$, et cetera. Throughout we assume that there are no ties at all among the observations (see Section 5), and that information on one group does not have any effect on probabilities of random quantities corresponding to the other group, so that X_{a,n_a+1} and X_{b,n_b+1} are independent and that data from group A does not influence our probabilities for X_{b,n_b+1}, and vice versa. We summarize this by stating that the groups are independent.

We require some additional notation, effectively counting the number of observed event times from group B to the left of observations from group A:

$$\begin{aligned} s_b(t_{a,i}) &= \#\{t_{b,j} \,|\, t_{b,j} < t_{a,i},\ j = 1, \ldots, u_b\}, \\ s_b(c^i_{a,k}) &= \#\{t_{b,j} \,|\, t_{b,j} < c^i_{a,k},\ j = 1, \ldots, u_b\}, \end{aligned}$$

for $i = 1, \ldots, u_a$ and $k = 1, \ldots, l_{a,i}$. Similarly, we need notation for the number of right-censoring times from group B in the interval $(t_{b,s_b(t_{a,i})}, t_{a,i})$:

$$s^c_b(t_{a,i}) = \#\{c_{b,j} \,|\, c_{b,j} \in (t_{b,s_b(t_{a,i})}, t_{a,i}),\ j = 1, \ldots, u_b\},$$

for $i = 1, \ldots, u_a + 1$.

The main results of this paper, namely the lower and upper probabilities for events $X_{a,n_a+1} > X_{b,n_b+1}$, based on the assumptions rc-$A_{(n_a)}$ and rc-$A_{(n_b)}$, are presented as a theorem below. The proof of the theorem is simplified via a lemma, which we present first, and which justifies the use of a variety of the theorem of total probability with conditioning on nested intervals, with probability distributions partially specified via M-function values.

Lemma 1 *For $s \geq 2$, let $J_l = (j_l, r)$, with $j_1 < j_2 < \ldots < j_s < r$, so we have nested intervals $J_1 \supset J_2 \supset \ldots \supset J_s$ with the same right end-point r (which may be infinity). We consider two independent real-valued random quantities, say X and Y. Let the probability distribution for X be partically specified via M-function values, with all probability mass $P(X \in J_1)$ described by the s M-function values $M_X(J_l)$, so $\sum_{l=1}^{s} M_X(J_l) = P(X \in J_1)$. Then, without additional assumptions, we have*

$$\sum_{l=1}^{s} P(Y < j_l) M_X(J_l) \leq P(Y < X, X \in J_1) \leq P(Y < r) P(X \in J_1),$$

and these bounds are optimal, so they are the maximum lower and minimum upper bounds that generally hold.

Proof. For any number s of nested intervals, the proof follows the same principle, so for ease of notation we present it for $s = 3$. We use the theorem of total probability to condition further on the partition $\{J_3, J_2 \setminus J_3, J_1 \setminus J_2\}$ of J_1 for the random quantity X. The probability distribution of X on J_1 is partially specified via M-function values for X defined on J_1, J_2, J_3. Let $M_X^l(J)$ denote the (unknown) part of the M-function value $M_X(J_l)$ that is actually in $J \subset J_l$, so we have

$$\begin{aligned}
P(X \in J_3) &= M_X^3(J_3) + M_X^2(J_3) + M_X^1(J_3), \\
P(X \in J_2 \setminus J_3) &= M_X^2(J_2 \setminus J_3) + M_X^1(J_2 \setminus J_3), \\
P(X \in J_1 \setminus J_2) &= M_X^1(J_1 \setminus J_2), \\
M_X(J_1) &= M_X^1(J_1 \setminus J_2) + M_X^1(J_2 \setminus J_3) + M_X^1(J_3), \\
M_X(J_2) &= M_X^2(J_2 \setminus J_3) + M_X^2(J_3), \\
M_X(J_3) &= M_X^3(J_3).
\end{aligned}$$

These M-function values are not further specified, but we can now use the theorem of total probability, and then derive bounds by solving the constrained optimiza-

tion problems. The lower bound follows from (with $J_4 = \emptyset$ for ease of notation)

$$\begin{aligned} P(Y < X, X \in J_1) &= \sum_{l=1}^{3} P(Y < X, X \in J_l \setminus J_{l+1}) \\ &= \sum_{l=1}^{3} P(Y < X \,|\, X \in J_l \setminus J_{l+1}) P(X \in J_l \setminus J_{l+1}) \\ &= P(Y < X \,|\, X \in J_1 \setminus J_2) M_X^1(J_1 \setminus J_2) + \\ &\quad P(Y < X \,|\, X \in J_2 \setminus J_3)[M_X^2(J_2 \setminus J_3) + M_X^1(J_2 \setminus J_3)] + \\ &\quad P(Y < X \,|\, X \in J_3)[M_X^3(J_3) + M_X^2(J_3) + M_X^1(J_3)]. \end{aligned}$$

With the constraints on these M-function values as given above, the lower bound is achieved by effectively putting the probability masses for X at the infimums of the intervals on which they are defined, so setting

$$M_X^1(J_2 \setminus J_3) = M_X^1(J_3) = M_X^2(J_3) = 0,$$

and taking the lower bounds for the conditional probabilities for $Y < X$, given $X \in I$, for the relevant I above, by replacing $X \in I$ by $X = \inf(I)$, leading to the terms $Y < j_l$ in the lower bound. The upper bound can be derived simultaneously, but is rather trivial as these nested intervals have the same right end-point. The fact that these bounds are optimal, without additional assumptions, follows easily from this construction. □

Bounds for the probability of $X_{a,n_a+1} > X_{b,n_b+1}$, based on rc-$A_{(n_a)}$ and rc-$A_{(n_b)}$, are presented in the following theorem. As these bounds are optimal, without any additional assumptions, they are lower and upper probabilities [15], which we denote by $\underline{P}(X_{a,n_a+1} > X_{b,n_b+1})$ and $\overline{P}(X_{a,n_a+1} > X_{b,n_b+1})$, respectively.

Theorem 1 *Assume that data are available from two independent groups, A and B, following the notation presented above. Based on the assumptions rc-$A_{(n_a)}$ and rc-$A_{(n_b)}$, predictive comparison of these two groups can be based on the following lower and upper probabilities for $X_{a,n_a+1} > X_{b,n_b+1}$,*

$$
\begin{aligned}
& \underline{P}(X_{a,n_a+1} > X_{n_b+1}) \\
= & \sum_{i=0}^{u_a} \left\{ \left[\sum_{j=0}^{s_b(t_{a,i})-1} P(X_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \right] M_{X_{a,n_a+1}}(t_{a,i}, t_{a,i+1}) \right. \\
& \left. + \sum_{k=1}^{l_{a,i}} \left(\left[\sum_{j=0}^{s_b(c_{a,k}^i)} P(X_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \right] M_{X_{a,n_a+1}}(c_{a,k}^i, t_{a,i+1}) \right) \right\}, \\
& \overline{P}(X_{a,n_a+1} > X_{b,n_b+1}) \\
= & \sum_{i=0}^{u_a} \left\{ \left[\sum_{j=0}^{s_b(t_{a,i+1})-1} P(X_{b,n_b+1} \in (t_{b,j}, t_{b,j+1})) \right. \right. \\
& + P(X_{b,n_b+1} \in (t_{b,s_b(t_{a,i+1})-1}, t_{b,s_b(t_{a,i+1})})) \\
& \left. \left. + \sum_{l=1}^{s_b^c(t_{a,i+1})} M_{X_{b,n_b+1}}(c^{s_b^c(t_{a,i+1})}, t_{b,s_b(t_{a,i+1})+1}) \right] P(X_{a,n_a+1} \in (t_{a,i}, t_{a,i+1})) \right\}.
\end{aligned}
$$

Proof. These lower and upper probabilities are derived by first writing

$$
P(X_{a,n_a+1} > X_{b,n_b+1}) = \sum_{i=0}^{u_a} P(X_{b,n_b+1} < X_{a,n_a+1}, X_{a,n_a+1} \in (t_{a,i}, t_{a,i+1})),
$$

and then applying the above lemma for each of the terms within this sum, and using the intervals on which the M-function values for X_{a,n_a+1} are defined according to rc-$A_{(n_a)}$. Then, bounds for the resulting probabilities (compare the lemma above) for X_{b,n_b+1} are determined, based on the corresponding M-function values according to rc-$A_{(n_b)}$, where a lower bound is derived by including only the M-function values on intervals that are fully included in the interval in the event of interest, and the upper bound is derived by including all M-function values on intervals that have non-empty intersection with the interval in the event of interest. Further details are relatively straightforward (see Yan [16] for a complete proof). □

These lower and upper probabilities are not available in a nice closed form. However, calculation is relatively easy as the individual terms are all product forms following from the definition of rc-$A_{(n)}$. If the data do not include any right-censorings, these lower and upper probabilities are identical to those presented by Coolen [3]. Although these formulae become fairly complex, the underlying idea for these optimal bounds is straightforward. The lower probability for $X_{a,n_a+1} > X_{b,n_b+1}$, based on the rc-$A_{(n)}$ assumptions per group, puts the probability masses as specified by the M-function values for X_{a,n_a+1} at the infimums of the intervals on which corresponding M-function values are specified, and for X_{b,n_b+1} at the supremums of the intervals, so at this bound the probability masses

are effectively least supportive for this event, given the partial specifications via M-function values. Of course, the upper probability just relates to these probability masses being put at the other end-points per interval.

We have presented the lower and upper probabilities for $X_{a,n_a+1} > X_{b,n_b+1}$. Similar results are available for the complementary event $X_{b,n_b+1} > X_{a,n_a+1}$, which can be derived by interchanging the indices for the groups above. However, it is not necessary to calculate lower and upper probabilities for both these events, because the well-known conjugacy property for imprecise probabilities [15], $\underline{P}(E) = 1 - \overline{P}(E^c)$, holds, where E^c is the complementary event of E. Informally, this holds because our bounds are optimal, and correspond to the same assessments based on the rc-$A_{(n)}$ assumptions per group. Alternatively, one could only compute either the lower or upper probabilities for both these events, requiring only a single algorithm, and using this relation to derive the other imprecise probabilities of interest.

Implicit in our results is that the probability of $X_{a,n_a+1} = X_{b,n_b+1}$ is zero, which is reasonable for our method as long as there are no ties among the event times of different groups (it would become a problem if a particular event time had been observed twice or more in each group, we discuss ties briefly in Section 5), and which is a consequence of our method of comparison, where effectively we always put probability masses at end-points of different intervals. It should be remarked, however, that a positive upper probability for $X_{a,n_a+1} = X_{b,n_b+1}$ could also be justified on the basis of these rc-$A_{(n)}$ assumptions, but doing so consistently would have made the analysis presented here more awkward, with little relevance for most practical situations.

4 Examples

We illustrate our nonparametric predictive method for comparison of two groups of lifetime data via two examples. We also compare our method with Mantel's two-sample test for censored data (see Section 11.7 of Hollander and Wolfe [10] for details), an established nonparametric method for such comparison, and discuss the important difference between our predictive approach and Mantel's hypothesis test.

Example 2
The data for this example are given in Table 2, and were also used by Parmar and Machin [14] to illustrate nonparametric methods for survival data. It is a subset of data obtained from 183 patients entered into a randomised Phase III trial conducted by the Medical Research Council Working Party on Advanced Carcinoma of the Cervix.

Table 2: Cervical cancer survival data ($>t$: right-censoring at t).

Control (A)	New (B)
90	272
142	362
150	373
269	>383
291	>519
>468	>563
680	>650
837	827
>890	>919
1037	>978
>1090	>1100
>1113	1307
>1153	>1360
1297	>1476
1429	
>1577	

The data are on survival of 30 patients with cervical cancer, recruited to a randomised trial aimed at analysing the effect of addition of a radiosensitiser to radiotherapy ('*new* treatment', B), via comparison to the use of radiotherapy alone ('*control* treatment', A). Of these 30 patients, $n_a = 16$ received the control treatment A, and $n_b = 14$ received the new treatment B. The data are in days since start of the study, the event of interest is death of the patient caused by this cancer. Further variables recorded for patients in the original study are not taken into account (see Parmar and Machin [14] for further references to the original study), we only use this subset of all the data to illustrate our new method for comparison of two such groups of data.

Using the method presented in Section 3, we compare these two groups of data predictively, by focussing on future observations $X_{a,17}$, assuming rc-$A_{(16)}$, and $X_{b,15}$, assuming rc-$A_{(14)}$. The corresponding lower and upper probabilities are

$$\underline{P}(X_{a,17} > X_{b,15}) = 0.226 \text{ and } \overline{P}(X_{a,17} > X_{b,15}) = 0.473,$$

which, by the conjugacy property for imprecise probability, imply

$$\underline{P}(X_{b,15} > X_{a,17}) = 0.527 \text{ and } \overline{P}(X_{b,15} > X_{a,17}) = 0.774.$$

These imprecise probabilities indicate that a preference for the new treatment B over the control treatment A would be reasonable, if no further information (e.g. on side-effects) is taken into account, and if one aims at surviving longer. In particular from an individual's perspective, this seems to be a natural inference if choice between two treatments is possible.

Although we do not discuss it explicitly here, such a choice could also take further aspects into account via our general rc-$A_{(n)}$-based inferential method. For example, a patient may prefer the treatment with maximum lower probability of surviving a particular length of time, it is fairly straightforward to calculate such lower probabilities per treatment in our approach [6].

From a classical nonparametric point of view, inference on the difference between survival chances for the two treatments could, for example, be based on application of Mantel's two-sample test for censored data, which is a rank-based test of a null-hypothesis of two equal survival functions, using asymptotic normality of the relevant test statistic. Applying this test for these cervical cancer survival data leads to a one-sided p-value of 0.1020, which may not be regarded as strong enough evidence against the null-hypothesis.

Example 3

The data for this example are given in Table 3, and were also used by Hollander and Wolfe [10] to illustrate Mantel's test. These data are from a clinical trial on Hodgkin's disease, a cancer of the lymph system. Two treatments were considered, a radiation treatment of the affected node (Treatment A; 25 patients), and a radiation treatment of the affected node plus all nodes in the trunk of the body (Treatment B; 24 patients). The data represent the relapse-free survival times in days. If a relapse had not occurred before the end of the study, then the observation for that patient is right-censored.

Our method, as presented in Section 3, applied to these data, leads to predictive imprecise probabilities

$$\underline{P}(X_{b,25} > X_{a,26}) = 0.557 \text{ and } \overline{P}(X_{b,25} > X_{a,26}) = 0.893.$$

These values indicate that the data suggest pretty strongly that $T_{b,25} > T_{a,26}$, hence it seems to be in a patient's best interest to opt for Treatment B. Applying Mantel's test to these data leads to an approximate one-sided p-value of 0.0006, which suggests very strongly that the survival functions corresponding to these two treatments are not equal.

Table 3: Hodgkin's disease survival data ($>t$: right-censoring at t).

Treatment A		Treatment B	
86	822	173	>1726
107	836	498	>1763
141	>1309	615	>1807
296	1375	950	>1879
312	>1378	>1190	>1889
330	>1446	>1242	>1897
346	>1540	1408	>1968
364	>1645	>1493	>1972
401	>1818	>1572	>2022
419	>1910	>1576	>2070
505	>1953	>1585	>2177
570	>2052	>1684	
688		>1699	

Clearly, testing equality of survival functions is quite a different inference than our predictive comparison, and it is not unreasonable to consider the outcome of both when trying to get more insight into the different survival chances per treatment. In Example 2, our method suggests that the new treatment would be better for a future patient than the control treatment, although Mantel's test does not strongly reject the hypothesis that both survival functions could be equal. In Example 3, the conclusions from both methods seem to agree more.

In general, it could also happen that Mantel's test would reject the null hypothesis, while we would end up with lower and upper probabilities both close to 0.5, so care should be taken on interpretation of the results of our method and Mantel's test. In situations where the real problem of interest is naturally in terms of comparison of next observations, we believe that our new method should be preferred.

The imprecision in our upper and lower probabilities in Examples 2 and 3 is not unreasonably large, in particular when considering the relatively large number of right-censored observations. This is explicitly due to our assumption rc-$A_{(n)}$, without this exchangeability-related assumption for the residual times till event for the right-censored items, logical bounds on the relevant conditional probabilities would be much wider.

5 Concluding remarks

We suggest that our new method for comparison of two groups of survival data is particularly useful in situations where such comparison takes place from a single

individual's perspective, e.g. when a person has a choice between the two treatments. If one has more relevant information, e.g. covariates or prior knowledge, some established statistical methods will be more appropriate. Our method can then still serve as a sort of base method, which can provide insight into the effect of further information or model assumptions, used with those alternative methods, by comparing the ultimate inferences. Extending our approach to possible inclusion of covariates is an interesting and relevant topic for future research.

Generalization of this approach to more than two groups of data is feasible, in a way similar to Coolen and van der Laan [5], who considered this problem without censored observations. It is also possible to extend attention to multiple future observations per group, but this would lead to rather complex computations due to dependence of such future observations for the same group [4, 7].

Throughout, we have assumed that there are no ties in the data. If there are ties, these can relatively easily be taken into account by breaking the ties, so assuming that tied values are only nearly identical, applying our method, and then letting the differences decrease to zero. For ties between the groups, one should break them into all possible orderings among the groups, calculate lower (upper) probabilities for each such ordering, and then take the minimum (maximum) of all these lower (upper) probabilities as the actual lower (upper) probability to be used for the comparison.

Acknowledgements

We are grateful to the referees for thoughtful comments and suggestions that have led to better presentation.

References

[1] T. Augustin, and F.P.A. Coolen. Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, to appear, 2003.

[2] L.M. Berliner, and B.M. Hill. Bayesian nonparametric survival analysis (with discussion). *Journal of the American Statistical Association*, 83:772-784, 1988.

[3] F.P.A. Coolen. Comparing two populations based on low stochastic structure assumptions. *Statistics & Probability Letters*, 29:297-305, 1996.

[4] F.P.A. Coolen. Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters*, 36:349-357, 1998.

[5] F.P.A. Coolen, and P. van der Laan. Imprecise predictive selection based on low structure assumptions. *Journal of Statistical Planning and Inference*, 98:185-203, 2001.

[6] F.P.A. Coolen, and K.J. Yan. Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, to appear, 2003.

[7] B.M. Hill. Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63:677-691, 1968.

[8] B.M. Hill. De Finetti's theorem, induction, and A_n, or Bayesian nonparametric predictive inference (with discussion). In *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, Eds. Oxford University Press, 211-241, 1988.

[9] B.M. Hill. Parametric models for $A_{(n)}$: splitting processes and mixtures. *Journal of the Royal Statistical Society B*, 55:423-433, 1993.

[10] M. Hollander, and D.A. Wolfe. *Nonparametric Statistical Methods* (2nd ed.). Wiley, New York, 1999.

[11] J.F. Lawless. *Statistical Models and Methods for Lifetime Data*. Wiley, New York, 1982.

[12] C.F. Manski. *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, Massachusetts, 1995.

[13] W.Q. Meeker, and L.A. Escobar. *Statistical Methods for Reliability Data*. Wiley, New York, 1998.

[14] M.K.B. Parmar, and D. Machin. *Survival Analysis: a Practical Approach*. Wiley, Chichester, 1995.

[15] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[16] K.J. Yan. *Nonparametric predictive inference with right-censored data*. PhD-thesis, University of Durham (available from the corresponding author), 2002.

Frank Coolen is Reader in Statistics at the Department of Mathematical Sciences, University of Durham, Durham 3LE, United Kingdom. E-mail: Frank.Coolen@durham.ac.uk

Ke-Jian Yan finished his PhD studies, under supervision of Frank Coolen, at the Department of Mathematical Sciences, University of Durham, in August 2002.