Some matched comparisons of two distributions of survival time

BY CHRISTIANA KARTSONAKI

Department of Oncology, University of Oxford, Oxford OX3 7DQ, UK
christiana.kartsonaki@oncology.ox.ac.uk

AND D. R. COX
Nuffield College, Oxford OX1 1NF, UK
david.cox@nuffield.ox.ac.uk

SUMMARY

A theoretical analysis is made of the properties of various methods for comparing two distributions of survival time. The results are intended primarily to guide the choice of method of analysis for such simple comparisons as of a treatment versus a control, but the main implications are fairly general, illustrating the performance of different models in a range of conditions. For most of the models there is a parameter specifying the comparison of interest and the Fisher information per observation can be calculated for that parameter. That provides a succinct basis for comparison. Two of the models are semi-parametric and the others are based on exponential distributions with or without extra sources of variability.

Some key words: Exponential distribution; Exponential family model; Fisher information; Matched pairs; Parameter orthogonality; Semiparametric formulation; Transformation model; Weibull distribution.

1. INTRODUCTION

There are various ways of comparing two distributions of survival times. A matched pair design illustrates in simple form the enhancement of the precision of the estimated contrast between two treatments achievable by comparing like with like. The statistical analysis is straightforward under the usual Gaussian assumptions, but less so otherwise for two rather different reasons. One is the sometimes controversial application of heavy conditioning to reduce the impact of between-pair variation; for the extreme case of binary outcomes, see McNemar (1947) and Cox (1958). The other aspect, in some ways more critical for interpretation, concerns the appropriate definition of the parameter of interest used for comparing the two treatments.

For survival times, Holt and Prentice (1974) applied the proportional hazards model to matched pairs data. Their analysis treated the pair effects as fixed nuisance parameters that are eliminated using a conditional likelihood for each pair. They also considered parametric models in which the survival times followed Weibull distributions with the same shape and a different scale parameter in each pair. Wild (1983) extended this in various directions. In a broad review of the analysis of survival data Breslow (1975) discussed the work of Holt and Prentice (1974), in particular the difficulties in dealing with censoring. Woolson and Lachenbruch (1980) gave extensions of the usual matched pairs location rank test statistics for censored data stressing the reliability of that approach on an assumption of symmetry of the distribution of the differences. Oakes (1981), in a wide-ranging discussion, pointed out in particular difficulties in correspond-

In the present paper we calculate for matched pair survival data the Fisher information about the between pair effect arising from a number of formulations. Mostly, but not entirely, the parameter defining that effect is the ratio of means of exponential distributions. The very simple explicit results show the effects on precision of changing the type of specification and as such may be a pointer also to behaviour in more complex situations.

The present paper is organized as follows. First we consider unmatched and matched semi-parametric proportional hazards models. Then we take a number of formulations in which the contrast between the distributions in the two groups is a simple scale change, that is captured by a ratio. Finally we consider a parametric formulation in terms of a difference of rates. In all cases we give, where feasible, a simple analytical result specifying the Fisher information for the parameter comparing the two groups.

For ease of interpretation and calculation we often parametrize the two distributions symmetrically. For example, in comparing two exponential distributions, instead of denoting the rate parameters for, say, control and treated groups, by \( \rho \) and \( \rho \theta \), with \( \theta \) the parameter of interest, we write instead for the two parameters \( \rho / \psi \) and \( \rho \psi \). The objective is to achieve orthogonality of parameters and, in particular, simplification in computing the relevant Fisher information. Finally, for ease of final interpretation, we use a log ratio, \( \omega = \log \theta = 2 \log \psi \).

2. **Semi-parametric proportional hazards model with and without matching**

Suppose we have \( n \) matched pairs with time-to-event outcomes \((t_{0j}, t_{1j}), j = 1, \ldots, n\), where 0 denotes the control group and 1 the treatment group. Consider first a model of one kind of extreme generality, having a different baseline hazard within each pair, representing a form of stratification. For the \( j \)th pair let the control group and treated hazards be respectively

\[
h_{0j}(t) = h_j(t), \quad h_{1j}(t) = h_j(t)e^\omega.
\]

(1)

Let \( X_j = 0 \) or 1, according as \( t_{0j} < t_{1j} \) or \( t_{0j} \geq t_{1j} \). Then so long as the \( h_j(t) \) are totally arbitrary all the information about \( \omega \) is contained in the \( X_j \) and

\[
P(X_j = 1) = e^\omega/(1 + e^\omega),
\]

so that formal inference about the parameter is essentially that for a simple binomial parameter. It follows on differentiating the log likelihood of the \( X_j \) and taking expectations that the Fisher information about \( \omega \) is

\[
i^{(1)}_{\omega\omega} = ne^\omega/(1 + e^\omega)^2.
\]

(3)

At the null hypothesis \( \omega = 0 \) this takes its maximum value of \( n/4 \).

Now suppose that the two samples are not matched and that a single arbitrary baseline hazard is taken for all individuals. That is (1) applies with \( h_j(t) = h(t) \) for all \( m = 2n \) individuals. A partial likelihood can be obtained as follow. Order the times in increasing order and write \( C_j = X_1 + \ldots + X_j \) for the number of treated individuals out of the first \( j \) failures. Then the
Matched pairs for survival times

conditional distribution referring to the next failure is
\[ P(X_{j+1} = 1 \mid C_j = c_j) = \frac{(n - c_j)e^{\omega}}{(n - c_j)e^{\omega} + (n - j + c_j)}. \] (4)

A partial log likelihood can now be formed by addition over \( j \). Minus its second derivative is
\[ \sum \frac{(n - c_j)(n - j + c_j)e^{\omega}}{(n - c_j)e^{\omega} + (n - j + c_j)^2}. \] (5)

Analytical evaluation of the expected value of this seems difficult except at \( \omega = 1 \) when the denominator simplifies and \( E(C_j) = j/2 \). Then we have that to an adequate approximation \( E(\Sigma C_j^2) = \{E(\Sigma C_j)^2\} \) and the expected information is again
\[ i(\omega) = n/4. \] (6)

Both these models are essentially unchanged by an arbitrary monotonic change in the time variable so that some loss of information is to be expected compared with, in particular, formulations that use parametric formulations of the distribution of survival time.

3. Some models directly based on the exponential distribution

3.1. Two independent samples

As a rather extreme contrast to the semiparametric formulations of the previous section we now consider some representations based directly on the exponential distribution. In this special case, although not in general, the ratio \( e^{\omega} \) of hazards is the inverse ratio of means. The discussion is much simplified by the invariance of \( \omega \) under scale changes of the data.

We start with the simplest situation with two independent samples each of size \( n \) both exponentially distributed, the treated group with rate parameter written, for symmetry, as \( \rho \psi \) and the control group with rate parameter \( \rho / \psi \). Write \( S_1 \) and \( S_0 \) for the two sample totals. Then the log likelihood is
\[ l = 2n \log \rho - \rho \psi S_1 - \rho \psi^{-1} S_0. \] (7)

The information matrix is diagonal and the information about \( \psi \) is
\[ i(\psi) = 2n/\psi^2. \] (8)

In terms of the log ratio of rates \( \omega = \log(\psi^2) \), we have
\[ i(\omega) = i(\psi)^2 (d\psi/d\omega)^2 = n/2. \] (9)

The information per observation is thus double that for the semi-parametric formulation.

A parallel discussion is possible replacing the exponential distribution by a Weibull distribution. That is, the density \( \rho e^{-\rho t} \) is replaced by \( \rho \gamma (\rho t)^{\gamma - 1} \exp \{- (\rho t)^\gamma \} \). Provided the index parameter, \( \gamma \), of the Weibull distribution is the same in the two groups, the two groups may be compared by the ratio of the parameters \( \rho \) or equivalently by the ratio of means. We write as before the two values of \( \rho \) in the form \( \rho \psi \) and \( \rho / \psi \). It can now be shown that locally near the null hypothesis \( \psi = 1 \) the parameter \( \psi \) is estimated orthogonally to \( (\rho, \gamma) \) and that the expected information about \( \psi \) is \( 2\gamma n \) so that locally the result (9) is modified by the factor \( \gamma \), expressing, for example, that the decreased relative dispersion of the Weibull distribution as \( \gamma \) increases results in an increase in the Fisher information per observation.
3.2. Right censoring

A frequent complication with survival data is right censoring, either by design or by the occurrence of “failures” of a type not of immediate concern. Subject to the strong assumption that the censoring is uninformative the log likelihood (??) is modified to

\[(n_{f1} + n_{f0}) \log \rho + (n_{f1} - n_{f0}) \log \psi - \rho \psi S_1 - \rho \psi^{-1} S_0, \quad (10)\]

where \(n_{f1}, n_{f2}\) are numbers of failures in the two groups.

If both groups are censored with probability \(\pi\), that is the probability of observing a failure is \(1 - \pi\), then a proportion \(\pi\) of the information is lost whereas if the censoring is confined to one group the proportional loss is \((1 - \pi)(1 - \pi/2)^{-1}\). The derivation of both results depends strongly on the special features of the exponential distribution.

3.3. Matched pairs

Next suppose the data are in matched pairs with a different value of the parameter \(\rho\) for each pair, \(\rho_j\), say, for the \(j\)th pair, assumed arbitrary; a constant value of \(\psi\) is postulated. It is essentially immaterial for the present calculation whether the \(\rho_j\) are arbitrary constants or are independent and identically distributed with an arbitrary distribution, a general random frailty model. Invariance considerations show that the information about \(\psi\) is contained in the pair-wise ratios \(y_j = t_{1j}/t_{0j}\), where \(t_{1j}\) and \(t_{0j}\) refer respectively to treated and control individuals in the \(j\)th pair. The probability density of \(Y_j\) is \(\psi^2/(1 + \psi^2 y_j)^2\). Minus the second derivative of the marginal log likelihood based on the \(y_j\) is

\[
\frac{2}{\psi^2} \Sigma \frac{1 + 4\psi^2 y_j - \psi^4 y_j^2}{(1 + \psi^2 y_j)^2}. \quad (11)
\]

Now take expectations, treating the corresponding \(Y_j\) as independent random variables, to obtain, after evaluating the resulting integral,

\[
i^{(4)}_{\psi\psi} = 4n/(3\psi^2) \quad (12)
\]

or in terms of \(\omega\)

\[
i^{(4)}_{\omega\omega} = n/3. \quad (13)
\]

3.4. A random effects formulation

As another somewhat less general possibility suppose that the \(\rho_j\) in the previous formulation are independently randomly distributed in the gamma distribution with density

\[
\alpha(\alpha \rho)^{\beta-1} e^{-\alpha \rho}/\Gamma(\beta), \quad (14)
\]

where \((\alpha, \beta)\) are unknown. If \(T\) is exponentially distributed with rate parameter \(\rho\) having the above gamma distribution the probability density of \(T\) is

\[
\beta \alpha^\beta / (\alpha + t)^{\beta+1}. \quad (15)
\]

For the comparison of two independent samples we assume all \(2n\) values of the rate distribution have this form with constant \(\beta\) and with \(\alpha\) replaced by \(\psi/\alpha\) and by \(\psi/\alpha\) for the two samples. The log likelihood is

\[
2n \log \beta + 2n \beta \log \alpha - (\beta + 1) \{\Sigma \log(\alpha \psi + t_{1j}) + \log(\alpha/\psi + t_{0j})\}. \quad (16)
\]

It can be shown that \(\psi\) is estimated orthogonally from \((\alpha, \beta)\) and that on differentiating twice with respect to \(\beta\) and taking expectations the Fisher information for \(\psi\) is \(2\beta/\{\psi^2(\beta + 2)\}\). Thus,
in terms of the log ratio $\omega$ and characterizing the variability of the gamma distribution not by $\beta$
but by its coefficient of variation $\gamma = 1/\sqrt{\beta}$, we have

$$i^{(5)}_{\omega\omega} = \frac{n}{2(1 + 2\gamma^2)}.$$  \hfill (17)

Thus for very large $\beta$, corresponding to the comparison of exponentially distributed variables,
we obtain $i^{(2)}_{\omega\omega}$, whereas for small $\beta$ the information becomes small because a large additional
component of variance is present on top of the variation implied by the exponential distribution.

When the data are in matched pairs, we assume that the underlying $\rho_j$ are the same for the two
members of a pair. The resulting contribution to the log likelihood is

$$\log \beta + \log(\beta + 1) + \beta \log \alpha - (\beta + 2) \log(\alpha + t_{j1}\psi + t_{j0}/\psi).$$ \hfill (18)

Estimation of $\psi$ is orthogonal to that of $(\alpha, \beta)$. The corresponding Fisher information may be
shown to be $1/\psi^2$ times its value when $\psi = 1$, that is when treated and control groups are identical,
and the Fisher information is then

$$E \left\{ \frac{2(\beta + 2)(\alpha T_0 + 2T_0 T_1)}{(\alpha + T_0 + T_1)^2} \right\}.$$ \hfill (19)

This may be evaluated by taking first expectations over $\rho$ and then over the distribution of $T_0, T_1$
marginalized over $\rho$ to give, after a lengthy calculation,

$$i^{(6)}_{\psi\psi} = \frac{n(2(\beta + 2)}{(\beta + 3)\psi^2}.$$ \hfill (20)

In terms of the fractional coefficient of variation of the gamma distribution, $\gamma$, and, as before
expressing the information in terms of the log ratio $\omega$, we have that

$$i^{(6)}_{\omega\omega} = \frac{1 + 2\gamma^2}{2(1 + 3\gamma^2)}.$$ \hfill (21)

For large $\gamma$ we recover $i^{(4)}$ and for zero $\gamma$ we recover $i^{(3)}$, corresponding respectively to arbitrary
variation in the rate parameter $\rho$ and to the comparison of two exponentially distributed samples,
that is with constant $\rho$.

3.5. A systematic representation

One often attractive possibility to avoid the introduction of the large number of nuisance pa-
rameters $\rho_j$ explicit in this last discussion is to suppose that the variation between pairs can be
explained by whole-pair explanatory variables $z_j$, for example that $\rho_j = \rho_0 \exp(\kappa^T z_j)$, where $\kappa$
is an unknown vector of regression parameters. Then it follows from the form of the mixed
second-order derivatives of the log likelihood that the parameters $\psi$ and $\beta$ are orthogonal, so
that (9) applies. If the explanatory variables are at an individual rather than a pair level the same
conclusion will hold only if the variables $z$ have the same mean in the two groups.

In outline there is up to a two-fold ratio between formulations in the Fisher information for
estimating the log ratio of rates, namely between $n/4$ and $n/2$. Estimates based on a parametric
random effects model yield the same range of values depending on the amount of underlying
variation in rate present.

4. A DIFFERENT EXPONENTIALLY-BASED FORMULATION

When the exponential density $pe^{-\rho_0}$ is written in exponential family form, the canonical pa-
rameter is the rate, $\rho$, so that the discussion in the previous section has been in terms of ratios
of canonical parameters. The previous results could have been obtained from exponential family properties, although in this instance use of the scale invariance properties of ratios is more direct and also emphasizes the usefulness of ratios as dimensionless summaries of comparative properties. The exponential family discussion is more direct in terms of differences of canonical parameters, broadly comparable to using an additive hazards model rather than a multiplicative one. There is also in this situation sometimes the interpretative advantage that statistically independent sources of failure contribute additively to the hazard.

We therefore give a brief discussion of the matched pair analysis in terms of representations linear in the rate parameter. In the notation of Section 3 and, writing the rate parameters in the jth pair as $\rho_j - \Delta$ and $\rho_j + \Delta$, we have that the log likelihood from pair $j$ is

$$2 \log \rho_j - \rho_j(t_{j1} + t_{j0}) - \Delta(t_{j1} - t_{j0}).$$

This leads to inference based on the conditional distribution of the difference $T_{j1} - T_{j0}$, or equivalently just $T_{j1}$, given the pairwise totals $T_{j1} + T_{j0} = t_{j}$. The marginal density of $T_j$ is obtained by noting that

$$E(e^{-qT_j}) = \frac{\rho_j^2 - \Delta^2}{(\rho_j + \Delta + q)(\rho_j - \Delta + q)}$$

so that, on splitting into partial fractions and inverting, we have that the density of $T_j$ is

$$(\rho_j^2 - \Delta^2)(e^{-(\rho_j - \Delta)t_{j}} - e^{-(\rho_j + \Delta)t_{j}})/(2\Delta)$$

with a reduced form at $\Delta = 0$. It follows that the conditional density of, say, $T_{j1}$ given $T_j = t_j$, is for $\Delta \neq 0$

$$\frac{2\Delta e^{2\Delta t_{j1}}}{1 - e^{-2\Delta t_{j}}}.$$  (25)

This is a truncated exponential distribution. If $\Delta = 0$ it takes the limiting form that $U_j = T_{j1}/t_{j}$ is uniformly distributed over $(0, 1)$.

By contrast if we assume all $\rho_j$ are equal the corresponding distribution is that of $S_1 = \Sigma T_{j1}$ given $S_1 = \Sigma(T_{j0} + T_{j1}) = s.$

Thus to test the null hypothesis that $\Delta = 0$ we take $(\bar{U} - n/2)/\sqrt{(n/12)}$, where $\bar{U} = \Sigma U_j/n$, as having very close to a standard normal distribution, the exact distribution being the convolution of $n$ unit rectangular distributions.

It can be shown that for both matched and unmatched analyses, whereas the estimation procedures for additive and multiplicative formulations are quite different, the tests of the null hypothesis of the identity of rates are the same, even though the derivation and distributional calculations are strikingly different.

5. Discussion

Even for the relatively simple situations discussed in the present paper there are many further possibilities that may arise. The exponential distribution may be replaced, for example by the Weibull distribution, and larger and situations more complex than paired comparisons may be involved. In all the models studied here the survivor times in the two groups are assumed mutually independent once a suitable specification is established. An interesting alternative approach (Oakes and Manatunga, 1992) specifies a bivariate extreme value for the paired observations with an unknown parameter representing the dependence. The resulting calculations are quite
Matched pairs for survival times

complicated and appear not to lead to the type of simple comparative conclusions obtained in the simpler contexts studied in the present paper.

REFERENCES


