Comparing Survival Curves Using an Easy to Interpret Statistic

Kenneth R. Hess

Department of Biostatistics
The University of Texas M.D. Anderson Cancer Center
P. O. Box 301402, Unit 1411
Houston, TX 77230-1402
khess@mdanderson.org
Abstract

I describe a statistic for comparing two survival curves that has a clear and obvious meaning and has a long history in biostatistics. Suppose we are comparing survival times associated with two treatments A and B. The statistic operates such that if it takes on the value 0.95, then the interpretation is that a randomly chosen patient treated with A has a 95% chance of surviving longer than a randomly chosen patient treated with B. This statistic was first described in the 1950's, and was generalized in the 1960's to work with right censored survival times. It is a useful and convenient measure for assessing differences between survival curves. Software for computing the statistic is readily available on the internet.
When comparing two survival distributions, it is customary to present a p-value for a statistical test of the hypothesis that the two curves are identical. However, such a p-value does not convey the magnitude of the difference between the curves. Furthermore, it is largely influenced by sample size. Sometimes the p-value is accompanied by an estimate of the hazard ratio. This ratio does convey the magnitude of the difference between the curves, but may be difficult to interpret. In what follows, I describe a statistic that has a clear and obvious meaning and has a long history in biostatistics.

Suppose we have survival times associated with treatment A that are represented by $x$ and survival times associated with treatment B that are represented by $y$. Now, let $Pr\{x > y\}$ represent the probability that a random subject from treatment A survives longer than a random subject from treatment B. It can be shown that $0 \leq Pr\{x>y\} \leq 1$ and that $Pr\{y>x\} = 1 - Pr\{x>y\}$. In the special case when the survival curves for A and B are identical, $Pr\{x>y\} = 0.5$. If there is complete separation between the survival time distributions such that all of the $y$ values are greater than all of the $x$ values, then $Pr\{x>y\} = 0$. If there is complete separation between the distributions such that all of the $x$ values are greater than all of the $y$ values, then $Pr\{x>y\} = 1$.

In their classic 1947 paper on comparing two samples, $x$ and $y$, Mann and Whitney [1] considered the statistic $U = \#$ of pairs $(x_j, y_k)$ such that $x_j > y_k$. If $m$ is the number of $x$'s and $n$ is the number of $y$'s, it can be shown that $V = U/m*n$ is a consistent, unbiased estimate of $Pr\{x > y\}$ [2]. Efron [3] generalized this statistic to the situation when both $x$ and $y$ are subject to right censoring. In addition, he provided a convenient computing formula for the statistic in the presence of censored data. In a modest simulation experiment, Koziol and Jia [4] showed that Efron's $V$ statistic was not affected by the censoring distributions for $x$ and $y$. They provide SAS and R code for computing Efron’s $V$ statistic as well as a bootstrapped standard error [5].

$Pr\{x>y\}$ is used in Bayesian statistics but here the distributions for $x$ and $y$ are assumed known [6]. In this case, there are several situations where $Pr\{x>y\}$ has a simple form. E.g., if $x$ is exponentially distributed with mean $a$ and $y$ is exponentially distributed with mean $b$, then $Pr\{x>y\} = a/(a+b)$. In general, if the distributions for $x$ and $y$ are known, then $Pr\{x>y\}$ can be computed using numerical integration [6]. In computing the $V$ statistic, we only assume that the distributions for $x$ and $y$ are continuous. In particular, the forms and parameters of the distributions are not assumed known.
In the special case when the x values represent marker values for a set of non-diseased individuals and the y values represent marker values for a set of disease individuals, Pr\{x>y\} is equivalent to the area under the receiver operating characteristic (ROC) curve constructed to relate the marker values to disease status [7]. This statistic is widely used in diagnostic testing to quantify the difference in marker values between subjects with and without disease.

Harrell et al [8] introduced an index which estimates the probability of concordance between predicted and observed responses. As Koziol and Jia [4] illustrate, the ideas behind the computation of Harrell’s concordance index can be used to compute a statistic for comparing survival curves which is similar in many respects to Pr\{x>y\}. However, the value of this statistic depends on the censoring distributions for x and y [4].

As with any statistic, V has statistical uncertainty associated with its estimation. One approach to quantify this uncertainty is to use bootstrap resampling in which repeated random samples are drawn with replacement from the data and the statistic computed on the resulting samples. The variability of the statistic over these samples can be used to estimate the variability of the statistic in the original samples. We can use this measure of variability to estimate approximate confidence intervals for the statistic.

To illustrate the use of the statistic Pr\{x>y\} for comparing survival curves, I use two published datasets. The first dataset is 76 patients with Ewing’s sarcoma who were treated at the NCI [9]. The comparison is between 45 patients with low serum lactic acid dehydrogenase (LDH) and 31 patients with high LDH (Figure 1). For these data V = 0.89 with approximate 95% confidence interval = (0.87, 0.91). Here the numeric value of the V statistic, 0.9, is consistent with the large separation in the survival curves. The second dataset is 89 patients with locally advanced nonresectable gastric carcinoma [10]. The comparison is between 44 patients treated with chemotherapy and 45 patients treated with chemotherapy plus radiation (Figure 2). For these data, V = 0.61 with approximate 95% confidence interval = (0.57, 0.65). Here the numeric value of the V statistic, 0.6, is consistent with the modest early separation of the survival curves. The interpretation of the V statistic is that a randomly chosen patient treated with chemoradiation will have about a 61% probability of surviving longer than a randomly chosen patient treated with chemotherapy alone. The hazard ratio comparing chemoradiation to chemotherapy alone is 0.89 with p = 0.60. Thus the V statistic better conveys the visual separation in the survival curves and has a simpler interpretation.
Pr{x>y} yields useful results when survival curves are divergent (curves are initially close and then separate later in time) or convergent (curves are initially separate and then come together), but crossing survival curves may present a problem. E.g., when curves cross near their median values with early and late separation, Pr{x>y} may be close to 0.5. In effect, the separation in the survival curves on the left of the crossing point cancels out the separation on the right of the crossing point. In addition, Pr{x>y} is not useful for comparing more than two survival curves (except in a pairwise fashion).
References

1. Mann HB and Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics 1947;18:50-60.
7. Bamber DC. The area above the ordinal dominance graph and the area below the receiver operating characteristic curve graph. J Math Psychol 1975; 12:387-412

Figure Legends

1. Ewing’s sarcoma: survival according to LDH status.
2. Gastric carcinoma: survival according to treatment.
Ewing Sarcoma Data; LDH

Disease-Free Survival

Time (years)

Low
High